Hermite-Gauss functions in the analysis of a category of semiconductor optical devices

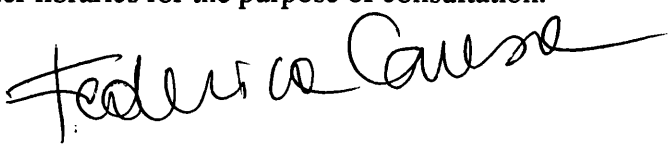
Causa, Federica

Award date:
1998

Awarding institution:
University of Bath

Hermite-Gauss Functions in the Analysis of a Category of Semiconductor Optical Devices

Submitted by Federica Causa
for the degree of
Doctor of Philosophy
of the University of Bath
1998



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

This thesis presents a novel self-consistent method for the analysis of active semiconductor devices, and in particular longitudinally non-uniform weakly guiding structures which have recently become of great interest.

The modelling of active semiconductor devices involves different aspects to be considered:

i) the solution of the wave equation for the analysis of the electromagnetic field in the device;

ii) the solution of the diffusion equation to obtain the distribution of the injected carriers;

iii) the description of the field-carrier interaction through a self-consistent scheme linking i) and ii), which produces the nonlinear effects typical of the operation of active devices. Note also that the longitudinal non-uniformity of the geometries of interest adds to the difficulties associated particularly with the solution of i) above since analytic solutions are generally not possible.

This thesis is composed of three main contributions reflecting the above three points. The attractive feature of the self-consistent analysis presented in this thesis is that it relies essentially on one method of solution, based on a function expansion procedure, which has been developed and successfully applied to solve both the field and the carrier diffusion equation. In this method the Hermite-Gauss functions form the very suitable, complete basis set in terms of which both the optical field and the carrier density profile may be accurately described. The numerical solution has been conveniently achieved with the collocation method, thus obtaining the compact Hermite-Gauss Collocation Method (HGCM).

The three stages of the work have been presented in sequence describing the development of the solution scheme and discussing the comparisons of the computed results with those from analytic or alternative numerical methods. In all cases such comparisons have been very satisfactory, demonstrating the versatility and accuracy of this compact HGCM scheme.

Acknowledgements

The dedicated and patient supervision of Dr J. Sarma has contributed enormously to this research work not only with interesting and illuminating discussions, but also with constructive criticism and continuous support.

Many thanks also to Prof. M. Milani and Prof. K. A. Shore for helpful discussions and support.

Stimulating inputs have also come from discussions with Prof. A. Sharma, Prof. A. K. Ghatak and with several other members of the Physics group of the Indian Institute of Technology, Delhi (India).

Thanks to the optoelectronics group of the University of Bath: S. Yunus, N. S. Brooks, T. Ryan, D. Hatten for continuous help and support; to Dr P. S. Spencer for helpful discussions and support; and to Dr R. Balasubramanyam for many useful discussions and support.

Thanks to Universita' degli Studi di Milano (Italy), to M.U.R.S.T. (Italy), to Alenia, Un' Azienda Finmeccanica S. p. A. (Italy) for financial support. Part of the research programme was financially supported via the University of Bath, which is also thanked for the facilities provided throughout.

Finally, many many thanks to my family for support and encouragement; and special thanks also to W. Hogarth, G. Arcimboldo, M. Schmelzer, J. B. Barrière, L. Stocks, W. Allen and S. D. Dog who made life much more enjoyable.

Contents

PAGE

Introduction vi

1 Review of some mathematical aspects of electromagnetic wave problems . 1

1.1 Maxwell's equations 2

1.1.1 Complex dielectric distribution 3

1.2 Interface conditions 4

1.3 The wave equation 5

1.3.1 The wave equation in homogeneous media 6

1.3.2 Vectorial wave equation for guided modes 7

1.4 Scalar wave equation 9

1.5 The radiation condition 10

1.6 Solutions of interest 10

1.6.1 Modal solutions 11

1.6.2 Quasi-modal solutions 11

1.7 The paraxial wave equation 12

1.7.1 Forward and reverse travelling waves 12

1.8 Power carried by electromagnetic waves 14

1.9 Closed waveguides 14

1.10 Open waveguides 16

1.10.1 The three-layer symmetric slab dielectric waveguide 17

Summary 21

Appendix 1.1 The effective dielectric constant method 23

Appendix 1.2 Plane wave decomposition and the paraxial wave equation . . 26

Appendix 1.3 Modes of parallel plate metal waveguides 29

Appendix 1.4 Functional form of TE and TM modes of the three-layer symmetric slab dielectric waveguide - scalar analysis . . . 32

References 36

2 The Hermite-Gauss functional form 38
2.1 Weak diffraction in homogeneous media: the Hermite-Gauss Beams . . 39
2.2 Modes of quadratic media: the Hermite-Gauss Eigenfunctions . . . 42
2.3 Approximation of the three-layer slab dielectric waveguide modes by the HGEs - the variational method 45
2.4 Connection between the Hermite-Gauss Eigenfunctions and the Hermite-Gauss Beams 47
Summary 48
Appendix 2.1 Derivation of the Hermite-Gauss Beams 50
References 52

3 Field analysis in optical dielectric structures: a brief review . . . 54
3.1 Longitudinally uniform structures: eigenvalue problem 55
3.1.1 The perturbation method 55
3.1.2 The global expansion method 57
3.1.3 The variational method 58
3.2 Longitudinally non-uniform structures: initial value problem . . . 59
3.2.1 The mode matching method 60
3.2.2 The beam propagation method 61
3.2.3 The total field propagation scheme 62
Summary 63
Appendix 3.1 First order perturbation technique 65
Appendix 3.2 BPM: derivation of the correction term e^{Γ} 68
References 70

4 Total field analysis using the Hermite-Gauss expansion method . . 73
4.1 Basis function set 74
4.2 Formalism for the HG expansion method 75
4.2.1 The eigenvalue problem 76
4.2.1.1 Complex media 77

4.2.2 The initial value problem 78
4.2.2.1 Paraxial wave equation 79
4.3 Discretisation 79
4.4 The collocation method 80
4.5 Choice of the width parameter (w_o) of the HG functions 82
Summary 83
Appendix 4.1 HG expansion method for the eigenvalue problem . . . 85
Appendix 4.2 HG expansion method for the initial value problem . . . 87
Appendix 4.3 Analytic method for the discretisation of the system of coupled differential equations (4.2.9) for the expansion coefficients $B_k(z)$ - derivation of equation (4.3.4) . . . 88
Appendix 4.4 The residual function 91
Appendix 4.5 The collocation method - formalism 92
References 95

5 Field analysis: results 96
5.1 Eigenvalue problem 97
5.1.1 Role of the w_o parameter in the HG function expansion. . . . 97
5.1.2 Piecewise constant refractive index distribution 98
5.1.3 Continuously varying refractive index distribution 99
5.1.4 Number of expansion terms, M 101
5.1.5 Waveguide bound modes - real dielectric distribution 104
5.1.6 Waveguide bound modes - complex dielectric distribution . . . 106
5.2 Initial value problem 108
5.2.1 Free-space propagation 108
5.2.2 Multimoded symmetric slab dielectric waveguide 109
5.2.3 Coupled waveguides 112
5.2.4 Longitudinal step discontinuities in waveguides 113
5.2.4.1 Approximate analysis of waveguide step discontinuities . . . 115
5.2.5 Field propagation in tapered dielectric structures 118

5.2.6 Parabolic structures 120
5.2.8 Sensitivity to w_o and N 124
Summary 127
Appendix 5.1 Variable transformation in the collocation method . . . 129
References 130

6 The carrier diffusion equation 131
6.1 Formulation 132
6.2 The HGCM for the solution of the diffusion equation 134
6.3 Results 136
6.3.1 Linear diffusion equation - comparison with the analytic solution . . 136
6.3.1.1 Unbounded region 136
6.3.1.2 Finite region 137
6.3.1.1 Finite region with mixed boundary conditions 138
6.3.2 Nonlinear diffusion equation - comparison with other methods of solution . 140
6.4 Sensitivity of the solution to M and w_o 142
Summary 146
Appendix 6.1 The analytic procedure to solve the carrier diffusion equation with the HG expansion method 147
Appendix 6.2 The collocation method for solving the carrier diffusion equation . 150
Appendix 6.3 Linear diffusion equation: analytic solution 151
References 153

7 The HGCM for the self-consistent solution of active semiconductor optical devices 154
7.1 Tapered laser devices 155
7.2 Active optical device modelling: the HGCM formulation. . . . 157
7.2.1 Fiel equation 158
7.2.2 Complex refractive index distribution 159
7.2.3 Carrier diffusion equation 160

7.2.4 Self-consistent method of solution 161

7.2.5 Numerical details 163

7.3 semiconductor amplifier and lasers: HGCM computed results . . . 165

7.3.1 Tapered geometry travelling-wave amplifier 165

7.3.1.1 Comparison with the mode matching technique 165

7.3.1.2 Forward and reverse travelling fields. 168

7.3.2 Linearly tapered laser 172

7. 3.2.1 $a_c = 0, \alpha = 0$ 172

7. 3.2.2 $a_c = 0, \alpha \neq 0$ 174

7. 3.2.3 $a_c \neq 0, \alpha \neq 0$ 176

7. 3.2.4 $a_c \neq 0, \alpha \neq 0, M = 51, w_o = 1.8\mu m$ 180

7. 3.2.5 $a_c \neq 0, \alpha \neq 0, M = 99, w_o = 1.3\mu m$, with deflectors 181

7.3.3 Discussion of the HGCM results for linearly tapered active devices . . 184

7.4 Parabolically tapered laser 185

7.5.1 Stripe laser 189

7.5.1 Numerical aspects of the solution for gain-guided devices . . . 193

7.6 Modelling of active tapered devices 195

7.6.1 Beam Propagation Method 195

7.6.2 Mode matching method 196

Summary 197

Appendix 7.1 Material parameters 199

Appendix 7.2 The average photon density 201

References 205

Future work 208

Conclusions 210

List of Tables 214

List of figures 215

Introduction

This thesis presents the outcome of research investigations which were motivated by the need to develop a convenient scheme for the modelling of a category of semiconductor optical devices. This need originates from the recent increasing interest in (optical) devices which are not uniform along the longitudinal axis. Obvious examples of such structures are tapered fibres for improving coupling to semiconductor optical sources. In fact, the complementary situation is also taking hold, namely that of specifically designing the source such that its output field profile matches as closely as possible with that of the fibre. Although optical sources are involved, the problem in those two examples are essentially related to passive devices. The recent, but rapidly growing interest in high power taper geometry semiconductor optical sources is an example of active devices with longitudinal non-uniformity. Reliable modelling of such devices is thus needed to design for improved and/or extended device characteristics.

The modelling of any such device is, broadly speaking, based on the analysis of two primary physical aspects - the electromagnetic (optical) field and the carrier (inversion population) distribution. The same is done for conventional (uniform) devices, but, now, the longitudinal non-uniformity introduces further complications. For example, whereas mode solutions are permissible for the uniform devices, no such convenient set is available, in general, for non-uniform structures. In the matter of carrier distribution also, the complexity of the solution is similarly increased. Consequently, it is important to develop a simple yet efficient modelling scheme. There are numerical methods available such as Finite Difference, Beam Propagation Method (BPM), etc., but they are dominantly numerical approaches. One of the motivations for the present research programme has been to develop a computational scheme which, by utilising the known properties of pertinent simpler structures, produces effective and efficient analyses.

Although the development of the new computational procedure is based on the motivation to model semiconductor optical devices, it is expected that the method will find applications also in other areas, such as in fluid dynamics and acoustic wave

propagation, since the basis of the work here is to solve for wave propagation in weakly non-uniform geometries.

Particularly for the modelling of weakly non-uniform structures the most immediate approach is that of the perturbation method, which relies on the known solution of the corresponding unperturbed (uniform) structure for describing the 'new' solutions to the problem. This approach is frequently and effectively used in (closed) metal guides, but the application to (open) dielectric structures creates considerable computational complications because of the need to include in the modal representation the radiation modes of the unperturbed (local) dielectric waveguide. However, the radiation modes are in integral form and it is this that creates the difficulty. Hence, other suitable expansion functions are often sought.

There are any number of function sets which are orthogonal and complete and, in principle, may be used instead of the (local) mode function set. However, for effective and efficient computation, the choice of the basis function set should satisfy the additional condition of not having an integral form, and also should have similarity with the typical field solutions for the structure to be modelled.

With this in mind the Hermite-Gauss Beam (HGB) set was first considered, since they provide an infinite but discrete set of orthogonal functions, which have close resemblance to the (diffracting) field solutions to be expected in weakly non-uniform structures. However, as discussed in this thesis, the HGBs are not amenable for the analysis of inhomogeneous media.

It was thus decided to consider the Hermite-Gauss Eigenfunctions (HGE), which are quite different from the HGBs in that they represent the modes of longitudinally uniform waveguides with quadratically varying dielectric distribution. Although they do not individually represent diffracting fields, they nevertheless resemble the modes of typical slab dielectric waveguides. The longitudinal variation is, then, conveniently accounted for by expanding the total field in longitudinally non-uniform structures in terms of the complete HGE set, and by using longitudinally varying expansion coefficients. Also the HGEs form a complete and discrete set of functions so that the desired properties for efficient modelling are satisfied.

By enforcing that the above HGE representation of the total field solution satisfies the wave equation for the structure to be modelled yields a set of coupled differential

equations for the expansion coefficients. This formalism, however, generally involves integral terms to be solved which, however, should be avoided for computational efficiency. This is the reason why the collocation numerical method has been adopted to evaluate the expansion coefficients, with the advantages of avoiding integration and of leading to a matrix formulation. Hence, the overall scheme developed in this thesis is referred to as the Hermite-Gauss Collocation Method (HGCM), and it is demonstrated that it forms a very efficient and effective compact scheme for the computation of field propagation in weakly non-uniform structures.

Before achieving the overall, self-consistent model it was needed to evaluate the gain/loss via the carrier (inversion population) diffusion equation. Various methods have been employed previously to solve the nonlinear diffusion equation, such as finite difference, cascaded matrix, Jacobi tri-diagonal method. Recognising that solutions to simple diffusion equations are Gaussian functions, it seemed sensible to try solving the diffusion equation with the HGCM. As discussed in detail in the thesis this new approach has proved to be an unqualified success.

The last stage of the development requires combining the methods for solving the optical field and the carrier distribution in a self-consistent scheme. The advantage and attraction of the HGCM is even more apparent at this stage since the two main computation routines are almost identical. There was no basic difficulty in achieving the final HGCM programme and it has thus been possible to obtain self-consistent solutions for weakly non-uniform active devices. Thus, the original objective has been achieved.

The thesis is structured as follows:

Chapter 1: review of some mathematical aspects relevant to the analysis of field problems. The aim of this chapter is to introduce the formalism that will be followed throughout the thesis.

Chapter 2: introduction to the Hermite-Gauss functional form in the two different sets of functions: i) the Hermite-Gauss Beams which are the eigensolution of the paraxial wave equation in a homogeneous medium, and hence represent diffracting solutions; ii) the Hermite-Gauss Eigenfunctions which are the modes of longitudinally

uniform quadratic media. It is the latter set which will be used in the remainder of the thesis as the basis set.

Chapter 3: review of some of the methods used to solve for field problems, which are distinguished as i) eigenvalue; and ii) propagation (initial value) problems.

Chapter 4: the HGCM formalism is presented for the solution of the scalar wave equation in open structures.

Chapter 5: the HGCM is applied to the solution of a variety of field problems, including i) eigenvalue problems in waveguides characterised by arbitrary dielectric distributions, including complex dielectric distributions; ii) diffraction in homogeneous media; iii) propagation across dielectric step discontinuities; iv) propagation in coupled waveguides; v) propagation in tapered structures. Comparisons of the results obtained with the HGCM with other methods of solution are discussed, together with some aspects of numerical analysis.

Chapter 6: the HGCM formalism is applied to the solution of the carrier diffusion equation. Results pertaining to linear and nonlinear forms of the diffusion equation are presented, discussing the comparisons with analytic solutions and with numerical solutions, respectively.

Chapter 7: the comprehensive self-consistent HGCM method is presented and applied to the analysis of active optical devices. Progressive results that show the development of the self-consistent scheme are discussed. Linear and parabolic tapered devices have been considered; results for the latter case have been compared with those from another method of solution based on the local mode expansion method.

Chapter 1

Review of some mathematical aspects of electromagnetic wave problems

This chapter presents a format of the basic electromagnetic field equations which will be referred to throughout this thesis.

The first part of the chapter describes the development of the formalism that leads to the wave equation. Thus, Maxwell's equations and the interface conditions are summarised in Sections 1 and 2. The wave equation is derived in Section 3 and relevant properties of this equation for two specific cases, i.e., propagation in homogeneous media and in waveguide structures, are discussed. In Sections 5, 6 and 7 two important approximations to the wave equation, namely the scalar and the paraxial approximations, are formally defined and some pertinent points are discussed in detail as a means of gaining further familiarity with the substance of the approximations. A rapid review of the Poynting vector in Section 8 concludes the first part of the chapter.

The second part of the chapter is dedicated to the review of (longitudinally uniform) guiding structures. The extensive application of waveguides in communication systems is evident; the discussion here is to emphasise the modal solutions that these structures admit. Waveguides have two main classifications: i) closed, ii) open. Examples of i) are hollow metal guides where the electromagnetic field is entirely contained inside the closed region. Dielectric waveguides are an example of ii), where the electromagnetic field extends laterally out to infinity. Metal guides are impractical and inefficient beyond microwave frequencies, and hence at higher frequencies (including optical frequencies) dielectric waveguides are extensively used.

The appropriate boundary conditions associated with the wave equation specify the set of characteristic solutions that are pertinent to any particular waveguiding structure. The characteristic solutions are known as the (eigen)modes of the waveguide.

The most important properties of closed guides, specifically hollow metal guides, are reviewed in Section 9. In Section 10 open waveguides are considered, referring for simplicity to the case of the three-layer symmetric slab waveguide. The main object here is to illustrate the particular features of the solution spectra of the wave equation that occur as a consequence of the different boundary conditions associated with the open as compared to the closed waveguides.

1. Maxwell's equations

The equations that govern the theory of electromagnetic waves derive from Maxwell's equations which are summarised here for the case of harmonic time dependent fields. Consider thus all fields to be of the form $\underline{\tilde{F}}(\underline{s},t) = \underline{F}(\underline{s})g(t) = \underline{F}(\underline{s})e^{i\omega t}$ where $\omega > 0$ (in rad/s) is a real quantity defined as the angular frequency, $\underline{F}(\underline{s})$ the complex vector function and $\underline{s}$ the spatial co-ordinates.

Maxwell's equations are

$$\underline{\nabla} \wedge \underline{E} = -i\omega\underline{B} \tag{1.1.1}$$

$$\underline{\nabla} \cdot \underline{D} = \rho \tag{1.1.2}$$

$$\underline{\nabla} \wedge \underline{H} = i\omega\underline{D} + \underline{J}_{tot} \tag{1.1.3}$$

$$\underline{\nabla} \cdot \underline{B} = 0 \tag{1.1.4}$$

where $\underline{E}$ (in V/m) and $\underline{H}$ (in A/m) represent the field intensity vector for the electric and magnetic field respectively, $\underline{D}$ (in C/m^2) and $\underline{B}$ (in weber/m^2 = tesla) the flux density vector of the electric and magnetic field respectively ($\underline{D}$ is also referred to as electric displacement vector), ρ is the electric charge density (in C/m^3); $\underline{J}_{tot} = \underline{J}_{source} + \underline{J}_c$ is the total current density term (in A/m^2), composed of the source and the conduction current density terms. In linear[1], isotropic[2] media the vectors $\underline{E}$, $\underline{D}$, $\underline{H}$, $\underline{B}$ and $\underline{J}_c$ satisfy the following (linear) constitutive relations

$$\underline{D} = \varepsilon_o\varepsilon_r\underline{E} \tag{1.1.5}$$

$$\underline{B} = \mu_o\mu\underline{H} \tag{1.1.6}$$

$$\underline{J}_c = \sigma\underline{E} \tag{1.1.7}$$

[1] A medium is linear when the constitutive relations hold independent of the magnitude of the field.
[2] A medium is isotropic when the constitutive relations hold independent of the direction of the field.

where ε_o is the electric permittivity of vacuum[1], ε_r the (real, dimensionless) relative dielectric constant of the medium, μ_o the magnetic permeability of vacuum[2], μ the (dimensionless) relative magnetic permeability of the medium and σ (in $m^{-1}ohm^{-1}$) is the conductivity of the medium, [1], [2].

In this thesis most of the theory pertains to electromagnetic wave propagation at high optical frequencies in source-free regions. Hence, the $\underline{J}_{source}$ term will be ignored henceforth (it is noted however that spontaneous emission may be accounted for through this explicit term, if needed). Thus, in the remainder of the thesis non-ferromagnetic, source-free media are considered for which $\mu = 1$, $\rho = 0$, and $\underline{J}_{source} = 0$ in equations (1.1.1) - (1.1.4).

1.1 Complex dielectric distribution

In optical media it is convenient to account for (stimulated) gain or loss through the conductivity, σ, of the medium. A complex dielectric distribution, ε, may be defined by

$$\varepsilon = \varepsilon_r - i\frac{\sigma}{\omega\varepsilon_o} = \varepsilon_r + i\varepsilon_i \tag{1.1.8}$$

where ε_r is the (real) relative dielectric distribution of the medium. Hence equation (1.1.8) may be written in terms of the (complex) refractive index distribution as

$$\varepsilon = (n_r + in_i)^2 \tag{1.1.9}$$

where n_r and n_i are the real and imaginary part of the refractive index, respectively. Making use of equation (1.1.8), equation (1.1.3) may be written in the compact form

$$\underline{\nabla} \wedge \underline{H} = i\omega\varepsilon_o\varepsilon\underline{E} \tag{1.1.10}$$

where ε is the complex dielectric constant defined in (1.1.8). The advantage of defining a complex ε is that Maxwell's equations (1.1.1) - (1.1.4) have the same form independent of the nature of the medium (ranging from insulators to conductors).

The wave equation, which determines the evolution of electromagnetic fields, can be derived from equations (1.1.1) - (1.1.4). This will be discussed in Section 3 for source-free media. However, it is worth reviewing at this point the conditions to

[1] $\varepsilon_o = 8.854\ 10^{-12}$ farad/m
[2] $\mu_o = 4\pi\ 10^{-7}$ henry/m

which the electromagnetic field is subject when travelling across an abrupt interface separating piecewise homogeneous media.

2. Interface conditions

For field analysis in piecewise homogeneous media with an abrupt interface, the most convenient approach is to solve Maxwell's equations in each homogeneous region and then match the solutions at the points of material discontinuity (interfaces) to ensure a unique solution everywhere in space. Most commonly two basic types of interfaces need be considered, i) that between two dielectrics and ii) that between a dielectric and a metal. The interface conditions can be derived from the integral form of Maxwell's equation (1.1.1) - (1.1.4), [1], [3], [4], [5] and are as follows.

The first condition, that for the tangential component of the electric field, derives from Ampère's law (equivalent to equation (1.1.1)), and is

$$E_{t(1)} = E_{t(2)} \tag{1.2.1}$$

where $E_{t(1)}$ represents the component of the electric field in medium 1 which is tangential to the interface, and correspondingly, $E_{t(2)}$ is that for medium 2, Fig. 1.2.1.

The condition on the normal component of the electric field is best defined by the electric displacement vector, from Gauss' law (equivalent to equation (1.1.2)), and reads

$$D_{n(1)} - D_{n(2)} = \rho_s \tag{1.2.2}$$

where $D_{n(1)}$ is the component of the dielectric displacement vector in medium 1 normal to the interface, and similarly for $D_{n(2)}$; ρ_s is the surface charge density on the interface (in C/m^2).

From the above equations it follows that in the absence of surface charges, the tangential component of the electric field is continuous, while the normal component of the electric field changes according to (1.2.2), i.e., $\varepsilon_1 E_{n(1)} = \varepsilon_2 E_{n(2)}$, with ε_1 and ε_2 the dielectric constants in medium 1 and 2, respectively.

The following conditions apply to the magnetic field. From Faraday's law (equivalent to equation (1.1.3)), obtain

$$H_{t(1)} - H_{t(2)} = J_s \tag{1.2.3}$$

where $H_{t(1)}$ denotes the component of the magnetic field which is tangential to the surface in medium 1, and similarly for $H_{t(2)}$; J_s is the surface current density (in A/m).

The condition on the normal component of the magnetic field derives from Gauss' (magnetic) law (equivalent to equation (1.1.4)):

$$B_{n(1)} = B_{n(2)} \tag{1.2.4}$$

where $B_{n(1)}$ is the component of the magnetic flux density vector normal to the surface in medium 1, and similarly for $B_{n(2)}$.

In the case of a metal-dielectric interface the metal is approximated as a perfect conductor (up to microwave frequencies). Since the fields inside a perfect conductor are zero, the interface conditions, with reference to the field components in the dielectric, become

$$E_t = 0 \tag{1.2.5}$$

$$D_n = \rho_s \tag{1.2.6}$$

$$B_n = 0 \tag{1.2.7}$$

$$H_t = J_s \tag{1.2.8}$$

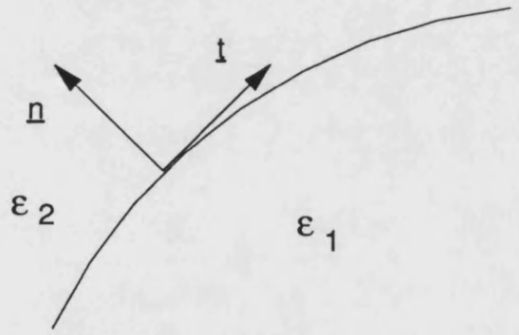


Fig. 1.2.1: Interface between two homogeneous media defined by ε_1 and ε_2; $\underline{n}$ and $\underline{t}$ are the unity vectors perpendicular and parallel to the interface, respectively.

3. The wave equation

Operating with $\nabla\wedge$ on equations (1.1.1) and (1.1.3) and using the vector identity $\Delta^2 \underline{v} = \underline{\nabla}(\underline{\nabla}\cdot\underline{v}) - \underline{\nabla}\wedge\underline{\nabla}\wedge\underline{v}$ [1] yields the vector wave equation

[1] This identity defines, in a general co-ordinate system, the operator Δ^2 which operates on the vector $\underline{v}$ = (v_x, v_y, v_z). Note that only in the Cartesian co-ordinate system $\left(\Delta^2 \underline{v}\right)_i = \underline{\nabla}\cdot\left(\underline{\nabla} v_i\right) = \nabla^2 v_i = \left(\partial_x^2 + \partial_y^2 + \partial_z^2\right) v_i$, with i = x, y, z.

$$\Delta^2 \underline{E} + k_o^2 \varepsilon \underline{E} = -\underline{\nabla}\left(\frac{\underline{\nabla}\varepsilon}{\varepsilon} \cdot \underline{E}\right) \tag{1.3.1}$$

where $k_o = \frac{\omega}{c} = \frac{2\pi}{\lambda_o}$ is the wave number in vacuum, with λ_o and $c = \frac{1}{\sqrt{\varepsilon_o \mu_o}}$ the wavelength and the velocity of light in vacuum, respectively, [5], [6]. Similarly, from equations (1.1.10) and (1.1.1) the corresponding wave equation for the magnetic field is

$$\Delta^2 \underline{H} + k_o^2 \varepsilon \underline{H} = -\frac{\underline{\nabla}\varepsilon}{\varepsilon} \wedge \underline{\nabla} \wedge \underline{H} \tag{1.3.2}$$

Equations (1.3.1) and (1.3.2) produce six scalar equations that determine the evolution of the six components of the electromagnetic field. All six field components are coupled through the right hand side term of equations (1.3.1) and (1.3.2), and by Maxwell's equations (1.1.1) - (1.1.4), [4].

3.1 The wave equation in homogeneous media

In a homogeneous medium the relative permittivity, ε, is independent of the spatial co-ordinates, hence the vectorial wave equations (1.3.1) - (1.3.2) become

$$\Delta^2 \underline{E} + k_o^2 \varepsilon \underline{E} = 0 \tag{1.3.3}$$

$$\Delta^2 \underline{H} + k_o^2 \varepsilon \underline{H} = 0 \tag{1.3.4}$$

In Cartesian co-ordinates equations (1.3.3) - (1.3.4) produce six scalar equations (of the same form), one for each field component:

$$\Delta^2 E_j + k_o^2 n^2 E_j = 0 \tag{1.3.5}$$

$$\Delta^2 H_j + k_o^2 n^2 H_j = 0 \tag{1.3.6}$$

with j = x, y, z. Each of the above six equations (1.3.5) - (1.3.6) is usually known as the scalar wave equation or Helmholtz equation.

A particular solution of (1.3.3) - (1.3.4) is the plane wave that depends only on one variable, e.g., z, and follows from assuming $\underline{E} = E_x(z)\hat{\underline{x}}$ in (1.3.3), where $\hat{\underline{x}}$ is the unity vector along the x-axis. Thus, the solution of the scalar equation for E_x is

$$E_x(z) = E_x^{(+)} e^{-kz} + E_x^{(-)} e^{+kz} \tag{1.3.7}$$

with $k^2 = -k_o^2 n^2$. The (constant) amplitudes $E_x^{(+)}$, $E_x^{(-)}$ are determined by the boundary conditions associated with the problem. Since, in general, the refractive index of a medium is complex, i.e., $n = n_r + in_i$, k in (1.3.7) is also complex, and hence

$$k = \alpha + i\beta \tag{1.3.8}$$

with α and β purely real. The parameter α is the attenuation (gain) constant (in neper/m). It can be shown that $n_i < 0$ corresponds to attenuating fields along the positive z direction, [7]. β in equation (1.3.8) specifies the magnitude of the phase in (1.3.7) and hence it is called phase constant (in radians/m).

The general vector solution of (1.3.3) is in this case

$$\underline{E}(z) = E_x^{(+)} e^{-\alpha z} e^{-i\beta z} \hat{\underline{x}} + E_x^{(-)} e^{+\alpha z} e^{+i\beta z} \hat{\underline{x}} \tag{1.3.9}$$

which may be seen as composed of two terms, the first a forward travelling wave, the second a reverse travelling wave. The phase velocities of the forward (+) and reverse (–) travelling waves are $v_{ph}^{(\pm)} = \pm\frac{\omega}{\beta}$, [7], [8].

For this plane wave with the electric field $\underline{E} = E_x(z)\hat{\underline{x}}$, the magnetic field can be derived from Maxwell's equation, and it can be shown that it has one field component in the y direction, $\underline{H} = H_y(z)\hat{\underline{y}}$, where $\hat{\underline{y}}$ is the unity vector parallel to the y-axis.

3.2 Vectorial wave equation for guided waves

In this section the vectorial wave equation is considered for the case of waveguides, that is, for dielectric distributions which are independent of the longitudinal variable z, i.e., $\varepsilon = \varepsilon(x, y)$. The purpose here is to highlight the importance of the $\nabla\varepsilon$ term in the vectorial wave equations (1.3.1) - (1.3.2) in the context of (longitudinally uniform) waveguides.

It is common to classify the waveguide solutions deriving from (1.3.1) - (1.3.2) into two categories: transverse electric (TE or H) waves for which the electric field has no component along the axis of propagation, and transverse magnetic (TM or E) waves for which the magnetic field has no component along the propagation axis. This distinction is valid in both closed and open waveguides and hence the following discussion is applicable generally to guiding structures. The TE and TM modes are particular solutions of the wave equation in waveguides. In the general case, all six

field components are present and the solutions are defined as hybrid waves, which may be viewed as TE and TM waves coupled by the boundaries imposed by the structure, [7], [9].

It is possible to prove that the electromagnetic field in the case analysed in this section may be fully represented by two field components, [5], [10]. Hence only the two transverse components of the electric field derived from equation (1.3.1) are considered here [a similar analysis can be applied to equation (1.3.2)].

Assume a mode solution described by the propagation factor $\exp(-i\beta z)$, and consider $\varepsilon = \varepsilon(x, y)$ in equation (1.3.1) in order to derive the equations for the two transverse components of the electric field, E_x and E_y:

$$\partial_x^2 E_x + \partial_y^2 E_x + \left(k_o^2 \varepsilon(x,y) - \beta^2\right) E_x + \partial_x \left[E_x \partial_x \ln(\varepsilon) + E_y \partial_y \ln(\varepsilon)\right] = 0 \quad (1.3.10)$$

$$\partial_x^2 E_y + \partial_y^2 E_y + \left(k_o^2 \varepsilon(x,y) - \beta^2\right) E_y + \partial_y \left[E_y \partial_y \ln(\varepsilon) + E_x \partial_x \ln(\varepsilon)\right] = 0 \quad (1.3.11)$$

with $\partial_x = \frac{\partial}{\partial x}$ etc. In the above equations three contributions may be highlighted:

1) the first three terms of equations (1.3.10) and (1.3.11) constitute the basic scalar wave equation for each individual field component. Thus, considering only the first three terms results in a simplified (degenerate) scalar analysis which is acceptable for weakly guiding waveguides, [4].

2) the term $\partial_x E_x \partial_x \ln(\varepsilon)$ in equation (1.3.10), [and respectively $\partial_y E_y \partial_y \ln(\varepsilon)$ in (1.3.11)] accounts for the polarisation effects on the field solutions (TE and TM modes). This term can be considered as a polarisation correction to the scalar wave equation, [11]. The propagation constant characterising the two polarisations (TE-TM) for the same mode number are different. However, without the polarisation correction the field solutions are degenerate, i.e., the modal profiles and corresponding propagation constants are identical for both polarisations. In the degenerate case it is sufficient to use the scalar wave equation to describe the chosen field component and derive the other field components from Maxwell's equations.

3) the last term $\partial_x E_y \partial_y \ln(\varepsilon)$ in (1.3.10) [and, correspondingly, $\partial_y E_x \partial_x \ln(\varepsilon)$ in (1.3.11)] represents the vectorial nature of the wave equation providing the coupling between the field transverse components. In the most general case the solution of the

(vectorial) wave equation involves all six components of the electromagnetic field (hybrid modes).

In weakly guiding structures it is common to consider the coupling terms between components to be weak so that the field solutions are assumed to be essentially linearly polarised. Since the cases analysed in this thesis predominantly involve weakly guiding structures all the corrections discussed above are neglected, i.e., the scalar analysis leads to sufficiently accurate results for the cases of interest.

4. Scalar wave equation

For the work described in this thesis the wave equation is solved mainly in the context of weakly guiding structures. In such cases the right hand side of equations (1.3.1) and (1.3.2) can be neglected, because either $\frac{\nabla\varepsilon}{\varepsilon} \approx 0$ (the rate of variation is sufficiently small) or $\underline{E} \perp \underline{\nabla}\varepsilon$. Hence, the wave equation becomes

$$\Delta^2\underline{E} + k_o^2\varepsilon(x,y,z)\underline{E} = 0 \qquad (1.4.1)$$

$$\Delta^2\underline{H} + k_o^2\varepsilon(x,y,z)\underline{H} = 0 \qquad (1.4.2)$$

The Cartesian co-ordinate system will be adopted in the following discussion because in this co-ordinate system the six scalar equations resulting from (1.4.1) - (1.4.2) have the same form:

$$\left[\nabla^2 + k_o^2\varepsilon(x,y,z)\right]F_j(x,y,z) = 0 \qquad (1.4.3)$$

where $\nabla^2 = \partial_x^2 + \partial_y^2 + \partial_z^2$, F_j is a component of either the electric or the magnetic field, j = x, y or z. In general only one of the above scalar (Helmholtz) equations is used to represent the chosen component of the electric or magnetic field, the other field components then follow from Maxwell's equations. The solution of equation (1.4.3) is uniquely determined once the associated boundary conditions imposed by the structure are specified.

In the remainder of the thesis weakly guiding structures will be considered in which the dominant guiding effects along the vertical (y) direction are accounted for separately, using the effective dielectric constant method (E. D. C.) [Appendix 1.1], or any other suitable method. Hence, use is made of the two-dimensional scalar wave equation in x and z,

$$\left[\partial_x^2 + \partial_z^2 + k_o^2 \varepsilon(x,z)\right] F(x,z) = 0 \qquad (1.4.4)$$

with F(x, z) the chosen field component and ε(x, z) the (effective) dielectric distribution in the (two-dimensional) medium.

5. The radiation condition

For the case of unbounded regions, such as open waveguides, the boundary condition associated with equation (1.4.4) is generally referred to as the Sommerfeld radiation condition. This condition imposes that if the sources are contained in a limited region, only outgoing waves are present at a large distance from that region, [4], [5], [12], [13].

The radiation condition assumes different expressions according to the co-ordinate system used and on the dimension of the physical space, [14]. For one-dimensional problems, in Cartesian co-ordinates the radiation condition in a lossless homogeneous medium is

$$\lim_{|x|\to\infty} \left(\partial_x F + ikF\right) = 0 \qquad (1.5.1)$$

where $k = k_o\sqrt{\varepsilon}$, with ε the dielectric constant of the medium, [14].

The general form of equation (1.5.1) is

$$\lim_{R\to\infty} R^{\frac{h-1}{2}} \left(\frac{\partial F}{\partial R} + ikF\right) = 0 \qquad (1.5.2)$$

where R is an independent variable in the direction of power flow, and h is the dimension of the physical space (in Cartesian co-ordinates $R^2 = x^2 + y^2 + z^2$), [15]. For lossy media the above conditions are simplified by assuming that the (total) electromagnetic field vanishes at infinity, [5], [13].

6. Solutions of interest

For the work discussed in this thesis it is assumed that the field solutions propagate predominantly along the longitudinal (z) direction, with the transverse profile either invariant (waveguide mode solutions) or slowly varying (weakly diffracting fields) with z. The main characteristics of these two types of solutions are presented below.

6.1 Modal solutions

For one-dimensional waveguides, with confinement along x and propagation axis along z, the method of separation of variables can be used to solve equation (1.4.4). Hence the field is written in the form

$$F(x,z) = F_m(x,z) = f_m(x)e^{\pm i\beta_m z} \tag{1.6.1}$$

which typically represents the waveguide modes. In (1.6.1) m is the mode number, $f_m(x)$ and β_m the field profile (eigenmode) and modal propagation constant (eigenvalue), respectively. For each mode, the eigenvalue β_m may be expressed in terms of an effective refractive index, n_m, defined by

$$n_m = \frac{\beta_m}{k_o} \tag{1.6.2}$$

6.2 Quasi-modal solutions

In the case of weakly diffracting fields it is advantageous to write the solution of equation (1.4.4) in the following form

$$F(x,z) = \begin{cases} f_+(x,z)e^{-ipz} \\ f_-(x,z)e^{+ipz} \end{cases} \tag{1.6.3}$$

where p is a suitably chosen (real positive) constant and $f_+(x,z)$ and $f_-(x,z)$ the field transverse profile of the forward and reverse travelling wave, respectively, since the harmonic time dependent factor $e^{+i\omega t}$ is assumed. For weakly guiding media, it can be assumed that $f_\pm(x, z)$ changes slowly during propagation, and, hence, for convenience, the large phase variation is separated out in the exponential term. The consideration of a slowly varying amplitude allows to introduce further simplifications to the scalar wave equation (1.4.4), namely the paraxial approximation which will be discussed in the next section.

The two types of solutions described in (1.6.1) and (1.6.3) are analysed in subsequent sections. In particular, the slowly varying field approximation presented in connection with equation (1.6.3) is extensively used in this thesis for the analysis of diffraction problems and for the study of field propagation in longitudinally non-uniform (dielectric) structures.

7. The paraxial wave equation

Consider in the following a weakly diffracting field, F(x,z), that is propagating essentially in the positive z-direction. Thus, from equation (1.6.3), the field is written as

$$F(x, z) = f_+(x, z)e^{-ipz} \tag{1.7.1}$$

In particular situations f(x, z) may be considered to be slowly changing with z. The analysis of electromagnetic fields with slowly varying amplitude is greatly simplified by the use of the paraxial approximation which is mathematically expressed by

$$\left|\partial_z^2 f(x,z)\right| << p\left|\partial_z f(x,z)\right| \tag{1.7.2}$$

and

$$\left|\partial_z^2 f(x,z)\right| << \left|\partial_x^2 f(x,z)\right| \tag{1.7.3}$$

Substituting (1.7.1) into the wave equation (1.4.4) obtain

$$\partial_x^2 f(x,z) + \partial_z^2 f(x,z) - p^2 f(x,z) - 2ip\partial_z f(x,z) + k_o^2 \varepsilon(x,z) f(x,z) = 0 \tag{1.7.4}$$

Use of (1.7.2) and (1.7.3) reduces equation (1.7.4) to the paraxial wave equation, [4], [16], [17],

$$\partial_x^2 f - 2ip\partial_z f + \left(k_o^2 \varepsilon(x,z) - p^2\right) f = 0 \tag{1.7.5}$$

The paraxial wave equation (1.7.5) has been used for most of the problems investigated in this thesis, particularly in the analysis of (longitudinally) weakly non-uniform structures defined by $\varepsilon = \varepsilon(x, z)$.

In the particular case of homogeneous media it is convenient to choose $p^2 = k_o^2 \varepsilon_h$ where ε_h is the dielectric constant of the medium, [Appendix 1.2]. Thus, equation (1.7.5) becomes

$$\partial_x^2 f - 2ip\partial_z f = 0 \tag{1.7.6}$$

7.1 Forward and reverse travelling waves

The use of equation (1.6.3) to describe the field implies that the direction of propagation has been chosen to be either in the positive or in the negative z-direction. For the general field solution of Helmholtz equation (1.4.4) both directions of propagation need to be considered. For example, in problems involving reflection of

the field at discontinuities the contribution of both the forward and the backward travelling fields would be necessary.

A general field may thus be written as

$$F(x, z) = f_+(x, z)e^{-ipz} + f_-(x, z)e^{+ipz} \tag{1.7.7}$$

with $f_+(x, z)$ and $f_-(x,z)$ depicting the forward (reverse) travelling wave. Substituting (1.7.7) into the wave equation (1.4.4) yields

$$\left(\partial_z^2 + \partial_x^2 - 2ip\partial_z + k_o^2\varepsilon - p^2\right)f_+e^{-ipz} + \left(\partial_z^2 + \partial_x^2 + 2ip\partial_z + k_o^2\varepsilon - p^2\right)f_-e^{+ipz} = 0 \tag{1.7.8}$$

However, the difficulty of solving equation (1.7.8) makes it necessary to introduce some approximations. Usually the two bracketed terms of equation (1.7.8) are solved separately, which is equivalent to assuming that the forward and reverse fields may be treated independently. Explicit boundaries along z are then satisfied by using the forward and reverse travelling waves.

In the case of weakly guiding structures, both $f_+(x, z)$ and $f_-(x, z)$ may be supposed to be slowly varying with z, and hence the analysis of the forward and, separately, of the reverse travelling wave may be simplified by the use of the paraxial approximation, i.e., for the forward

$$\left[\partial_x^2 - 2ip\partial_z + k_o^2\varepsilon(x,z)\right]f_+(x,z) = 0 \tag{1.7.9}$$

and for the reverse travelling wave

$$\left[\partial_x^2 + 2ip\partial_z + k_o^2\varepsilon(x,z)\right]f_-(x,z) = 0 \tag{1.7.10}$$

Although most of the structures analysed in this thesis are longitudinally non-uniform, which implies that the field is (continuously) reflected at the discontinuities, the continuously reflected field has been neglected in the analysis developed in this thesis. In those cases in which it is essential to consider also the reverse propagating field, e.g., in laser structures, use is made of the individual paraxial wave equations in both directions of propagation.

Consider the following initial value problem in which the initial field is given by $F(x, z = 0) = f_o(x)$. Assuming propagation along the positive z-direction in the range (0, L), the solution, $F(x, z = L) = f_L(x)$, is found solving equation (1.7.9). Note, however, that given the initial field $f_L(x) = F(x, z = L)$, equation (1.7.9) yields $F(x, z = 0) = f_o(x)$, as expected, while equation (1.7.10) produces a quite different result, $f_1(x) \neq f_o(x)$. This property is shown schematically in Fig. 1.7.1.

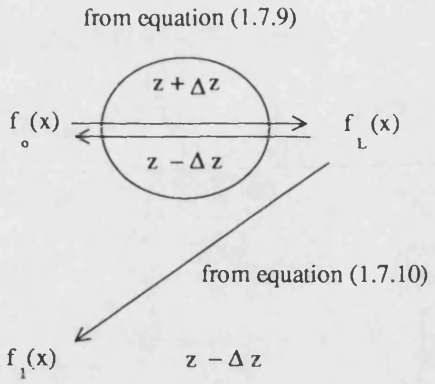


Fig. 1.7.1: Connections between the solutions of the paraxial wave equations (1.7.9) and (1.7.10).

8. Power carried by electromagnetic waves

The power carrier by an electromagnetic wave is determined by the Poynting vector, $\underline{S}$, which, for linear, isotropic media is defined by

$$\tilde{\underline{S}} = \tilde{\underline{E}} \wedge \tilde{\underline{H}} \tag{1.8.1}$$

where the symbol ~ indicates the space-time quantities. The Poynting vector may be interpreted as the power per unit area that flows at that point, [7]. At high frequencies the instantaneous power is very rapidly varying with time and hence it is common practice to use the time-averaged (complex) Poynting vector[1] which is called power density or irradiance (in W/m²). For harmonic time varying signals the time-averaged power density is defined as

$$\underline{S} = \tfrac{1}{2}\mathrm{Re}\left(\underline{E} \wedge \underline{H}^*\right) \tag{1.8.2}$$

The measurable quantity derived by the irradiance (1.8.2) is (the real part[2] of) the time-averaged power (P) flowing through a surface (A), i.e.,

$$P = \tfrac{1}{2}\mathrm{Re}\left[\int_A \left(\underline{E} \wedge \underline{H}^*\right) \cdot \underline{da}\right] \tag{1.8.3}$$

where $\underline{da}$ is the vector differential surface element.

9. Closed waveguides

Closed waveguides are characterised by the fact that the field is confined inside the (perfectly) reflecting boundaries. Consider in this section a parallel plate guide

[1] The time-average Poynting vector is derived from the Poynting theorem in the frequency domain, [3].

[2] The imaginary part of the complex Poynting power can be interpreted as reactive power, [7], [3].

formed by two perfectly conducting metal plates at $x = \pm a$, as shown in Fig. 1.9.1. TE and TM modes may be obtained in this case, as summarised in Appendix 1.3. However, in the following discussion the interest is in illustrating the relevant properties of the eigenvalues and eigenfunctions of the type (1.6.1) of closed waveguides, and hence the symbol $f_m(x)$ may represent either a TE or a TM mode.

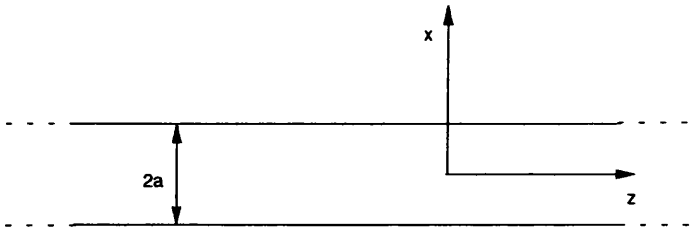


Fig. 1.9.1: Parallel plate metal closed waveguide: perfectly conducting boundaries at $x = \pm a$.

The modes of (longitudinally uniform) closed waveguides constitute an infinite complete set of discrete (bound) modes either propagating or evanescent. It can be demonstrated that the modes of any (lossless) waveguide are orthogonal to one another. This implies that each individual mode may propagate independently of the other modes, [9], [3], [18]. This property is expressed mathematically by the following integral:

$$\int_{-a}^{+a} f_m(x) f_n^*(x) dx = \delta_{mn} \tag{1.9.1}$$

where $f_n(x)$ and $f_m(x)$ are the modal fields. From the orthogonality property (1.9.1) also follows the power-orthogonality, [1], which means that if the waveguide supports more than one mode, the power associated with a general field composed of many modes of the guide is given by the summation of the contributions of the individual modes.

The orthogonality relation (1.9.1) is generally used for the normalisation of the mode fields, $f_m(x)$

$$\tilde{f}_m(x) = \frac{f_m(x)}{\sqrt{\int_{-a}^{+a} |f_m(x)|^2 dx}} \tag{1.9.2}$$

with $\tilde{f}_m(x)$ specifying the normalised m-th mode, [Appendix 1.3].

Another important property of the modes of closed waveguides is that of completeness, which permits the expansion of any field at any cross section of the waveguide in terms of the modal fields. That is, any total field, $F_o(x)$, inside the waveguide may be represented as

$$F_o(x) = \sum_{m=1}^{\infty} A_m \tilde{f}_m(x) \tag{1.9.3}$$

where $\tilde{f}_m(x)$ represent the modal fields, A_m the (constant) expansion coefficients. Using the orthogonality condition (1.9.1) the expansion coefficients in (1.9.3) may be determined by $A_m = \int_{-a}^{+a} F_o(x)\tilde{f}_m(x)dx$.

Since the modes propagate independently in the guide, the field at any longitudinal position, z, is given by

$$F(x,z) = \sum_{m=1}^{\infty} A_m \tilde{f}_m(x) e^{-i\beta_m z} \tag{1.9.4}$$

where each characteristic modal phase factor is specified by the corresponding propagation constant β_m. At any longitudinal position the power associated with the field expressed by (1.9.4) is

$$P = \sum_{m=1}^{\infty} P_m \propto \sum_{m=1}^{\infty} |A_m|^2 \int_{-a}^{+a} \left|\tilde{f}_m(x)\right|^2 dx \tag{1.9.5}$$

where P_m is the power of each mode. However, although the power obtained by the summation of the contribution of the individual modes is equal to the power of the incident field at all longitudinal positions, the profile of the field changes during propagation because of the different modal phase factors. The details necessary to calculate the power associated with TE/TM waveguides modes are summarised in Appendix 1.3.

10. Open waveguides

Metal plate waveguides are very effective at microwave frequencies because the high reflectivity of the boundaries confines the field inside the cavity. However, at higher frequencies these structures become too lossy and the alternative guiding mechanism of total internal reflection at dielectric interfaces is exploited to produce optical waveguides. In dielectric waveguides the wave is not totally confined inside the waveguide as in the case of closed guides, but extends out to infinity. The

achievement of waveguiding in dielectric waveguides made it possible to guide waves of higher frequency and thus to increase the capacity for information transmission.

In the following discussion reference is made to slab dielectric waveguides. Although they have some aspects in common with closed guides, open waveguides are characterised by the very distinctive feature of radiating into open space if any irregularity is present, [19]. Similarly to closed guides, open waveguides guide only a finite number of discrete modes, known as guided modes, or surface modes to underline the fact that the wave is guided by the dielectric surface. But, differently from closed waveguides, in open waveguides the complete spectrum of solutions involves also the category of radiation (quasi-surface) modes which form a continuous spectrum, [9], [19]. By requiring that the total field (instead of the individual modes) satisfies the radiation condition, a weaker condition is found by which the radiation modes become legitimate solutions of the wave equation, and hence the mode spectrum of open waveguides becomes complete, as discussed in detail in reference [19]. As in the case of closed guides, all modes of an open waveguide are mutually orthogonal in the integration over the plane transverse to the longitudinal axis. Further, any field in an open waveguide may be described in terms of the complete set of eigensolutions (guided modes and radiation modes) pertinent to the structure, [4]. This process of field expansion is similar to the one described in connection with closed waveguides, and is generally known as eigenmode expansion or local mode expansion. Deviations from perfect geometry not only convert power among guided modes of the dielectric waveguide, but also scatter power into the continuous spectrum of radiation modes, which is seen as radiation outside the dielectric waveguide, [4].

10.1 The three-layer symmetric slab dielectric waveguide

A symmetric three-layer slab waveguide is considered here with the purpose of briefly reviewing the characteristic features of the solutions of open waveguides (see e.g., [5], [20] for asymmetric open waveguides). Hence, assume $\partial_y = 0$.

The characteristic solutions of slab open waveguides may be classified as TE and TM modes. However, since each type of modes may be derived from one field component, in the following discussion it will be understood that for TE modes

$$F(x,z) = E_y(x,z) = f(x)e^{-i\beta z} \tag{1.10.1}$$

and $f(x) = e_y(x)$, while for TM modes

$$F(x,z) = H_y(x,z) = f(x)e^{-i\beta z} \tag{1.10.2}$$

and $f(x) = h_y(x)$. The other field components derive from Maxwell's equations (1.1.1) - (1.1.4), [Appendix 1.4]. Further, because of the symmetry of the dielectric distribution there are two separate sets modes, i.e., the symmetric and the anti-symmetric (TE/TM) modes, [Appendix 1.4]. This distinction will not be explicitly mentioned in the following discussion.

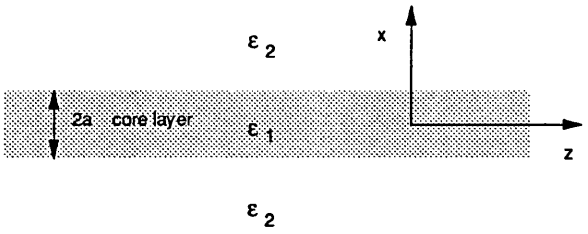


Fig. 1.10.1: Schematic of a longitudinally uniform three-layer symmetric slab waveguide, the core layer (shaded region) has a larger dielectric constant with respect to the cladding layers. The thickness of the waveguide core layer is 2a.

Consider thus a longitudinally uniform three-layer symmetric slab waveguide described by the dielectric distribution $\varepsilon = \varepsilon(x)$ independent of z. The layers are assumed to extend to infinity along the x and z axes with no variation along the y-direction. With reference to Fig. 1.10.1, $\varepsilon(x) = \begin{cases} \varepsilon_1 & |x| \le a \\ \varepsilon_2 & |x| \ge a \end{cases}$ with $\varepsilon_1 > \varepsilon_2$ (ε_1 and ε_2 real). Assuming a (bound) mode-solution of the type (1.6.1)

$$F(x, z) = f(x)e^{-i\beta z} \tag{1.6.1}$$

the wave equation (1.4.4) becomes

$$\frac{d^2 f_m(x)}{dx^2} + k_{x,m}^2 f_m(x) = 0 \tag{1.10.3}$$

where $k_{x,m}^2 = (k_o^2\varepsilon - \beta_m^2)$ determines the eigenvalue β_m corresponding to the eigensolution $f_m(x)$. Having assumed $\varepsilon(x)$ to be piecewise constant, equation (1.10.3) is solved separately in the three regions. The (unique) mode-solutions are obtained matching the solutions determined in each (homogeneous) region at the interfaces between the dielectrics.

The discrete (surface) modes are determined using the boundary (radiation) conditions pertinent to open structures, i.e., that the field decays to zero at infinity along the transverse axis:

$$|f_m(x)| \to 0 \qquad \text{for } |x| \to \infty \qquad (1.10.4)$$

The eigenmodes are characterised by a transverse distribution of the field, $f_m(x)$, that remains unchanged along the longitudinal direction z and by a propagation constant β_m. For each waveguide bound mode it is possible to define an effective refractive index $n_m = \frac{\beta_m}{k_o}$ [equation (1.6.2)]. The guided modes are descriptive of electromagnetic energy confined inside the guide; but to account for the energy radiating outside the guide the radiation modes need to be considered, [4]. The radiation modes are determined using the weaker condition that the field be finite at infinity

$$|f_m(x)| < \infty \qquad \text{for } |x| \to \infty \qquad (1.10.5)$$

instead of (1.10.4), and are characterised by having a continuum of values for the propagation constant (continuum spectrum). The expressions for the modal fields for guided and radiation modes are summarised in Appendix 1.4.

The overall spectrum of eigenvalues of equation (1.10.3) may be summarised in the graph of Fig. 1.10.2, [4], [20].

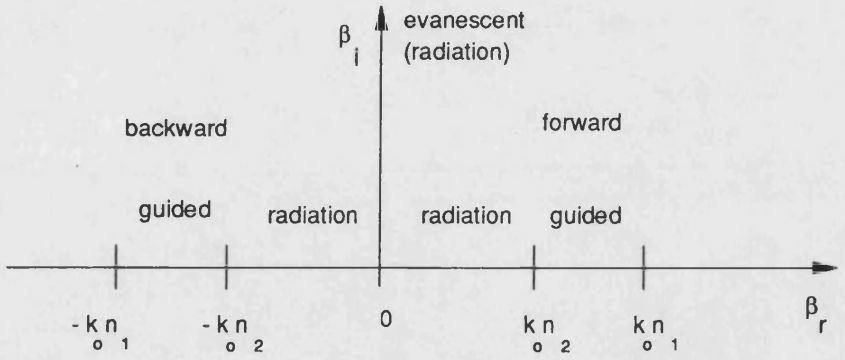


Fig. 1.10.2: Spectrum of the eigenvalues of an open waveguide (β_r and β_i refer to real and imaginary part of β respectively, n_1 is the highest value of the refractive index distribution, n_2 the lowest)

The solutions of the slab waveguide are usually classified as

i) $\beta > k_o n_1$ - the solutions exponentially grow in all three layers which implies infinite field amplitudes (and power) at large distances from the waveguide. This situation is not acceptable (physically impossible).

ii) $k_o n_1 > \beta > k_o n_2$ - a discrete number of bound (guided - surface) modes which vary sinusoidally inside the waveguide core and decay exponentially outside the waveguide; this set of β values forms the discrete spectrum of the waveguide.

iii) $k_o n_2 > \beta > 0$ - the solutions vary sinusoidally in all three layers: these are the radiation (quasi-surface) modes of the slab waveguide. Any value of β in this range is allowed, hence this set constitutes part of the continuum spectrum of the waveguide.

iv) $\beta < 0$ - the negative values of the propagation constant are an indication of the reverse direction of propagation, i.e., $\beta < 0$ represent backward travelling modes. The same distinctions as above apply in this region, with the signs appropriately changed.

v) imaginary β, i.e., $\beta = -i|\beta|$ with $0 < |\beta| < \infty$ - the corresponding eigensolutions are the evanescent (or reactive) modes. The inclusion of evanescent modes in field analysis is necessary for a detailed description of the field at the surface of the waveguide, but since they do not carry power away from the waveguide they can be neglected in the analysis of field propagation in dielectric waveguides, [4], [5]. The reactive modes correspond to the non-propagating modes below cut-off in closed waveguides, [5].

The cut-off condition for guided modes occurs when $\beta_m = k_o n_2$ and determines the frequency below which the mode ceases to propagate in the waveguide, i.e., $k_o a\sqrt{\varepsilon_1 - \varepsilon_2} = m\pi$, or, in terms of the cut-off frequency, $\omega_{c(m)}$:

$$\omega_{c(m)} = \frac{m\pi c}{a\sqrt{\varepsilon_1 - \varepsilon_2}} \tag{1.10.6}$$

with m = 0, 1, 2, ... Note that for symmetric (slab) waveguides the fundamental mode has no cut-off frequency, [4].

As already discussed for closed guides, the totality of the mode-functions $f_m(x)$ of open waveguides have the mathematical property of forming a complete set of

mutually orthogonal functions, [5]. For open waveguides the orthogonality condition may be written as

$$\int_{-\infty}^{+\infty} f_m(x) f_n^*(x) dx = 0 \qquad \text{for } m \neq n \tag{1.10.7}$$

Equation (1.10.7) holds for any combination of guided and radiation modes, [4]. The orthogonality relation (1.10.7) is typically used to normalise the mode functions. In particular, the normalisation condition for radiation modes involves the Dirac delta distribution, i.e.,

$$\frac{\beta}{\omega\mu_o} \int_{-\infty}^{+\infty} f(x,\rho) f^*(x,\rho') dx = P\delta(\rho - \rho') \tag{1.10.8}$$

here ρ represents the propagation constant, [Appendix 1.4] and P is the power flowing in the z-direction per unit length in the y-direction, [4]. Refer to Appendix 1.4 for the equations defining the power carried by waveguide TE/TM modes.

The totality of the modes of slab waveguides form a complete set, [4], and hence any electromagnetic field inside a dielectric slab may be expressed in the following form

$$f_{(y)}(x) = \sum_{\substack{m \\ \text{even} \\ \text{odd}}} c_m f_{(y)m}(x) + \sum_{\substack{\text{even} \\ \text{odd}}} \int_o^{+\infty} q(\rho) f_{(y)}(\rho) d\rho \tag{1.10.9}$$

Equation (1.10.9) holds for the y-component of the electric field of TE modes, $E_y(x,z) = f_y(x)e^{-i\beta z}$, the other components may be expanded similarly. The expansion coefficients of (1.10.9) are determined using the orthogonality condition (1.10.7). Although rigorously the complete set of modes ought to be used, as in (1.10.9), in many situations such as in weak dielectric step waveguides it is sufficient to consider only the set of the discrete modes to obtain satisfactory (approximate) results, which simplifies the analysis, [2], [4].

Summary

The fundamental equations for electromagnetic field analysis are summarised in the first part of this chapter. Particular attention is given to the implications of the scalar and paraxial approximations to the wave equation since they will be used substantially in the remainder of the thesis.

The second part of the chapter deals with the properties of the solutions of the wave equation. In particular, a basic comparison between closed and open guiding structures is discussed, the most important difference being the presence of a continuous spectrum in open waveguides.

Appendix 1.1

The effective dielectric constant method

Two dimensional, e.g., rectangular cross-section, dielectric waveguides are mathematically difficult to analyse since separation of variables is not applicable. For rectangular dielectric waveguides which produce mode profiles that have a large aspect ratio various approximate methods of analysis are possible. One of the simplest and, in that context, perhaps the best is the effective dielectric constant (E. D. C.) method, [22], [23]. For example, the structures of interest in this thesis are taken to be fabricated from typical multilayer semiconductor material, as shown in Fig. A1.1-1. The resulting three-dimensional structure has a dielectric distribution of the type $\varepsilon = \varepsilon(x, y)$. If the rib section is longitudinally non-uniform then the dielectric distribution is also z-dependent, i.e., $\varepsilon = \varepsilon(x, y, z)$. In such structures it can be convenient to reduce the problem to two dimensions, the transverse (x) and the longitudinal (z), using the E. D. C. method (along the vertical y-axis) to derive an effective dielectric distribution for a corresponding two-dimensional structure.

For the following derivation, [22], consider a waveguide of the type shown in Fig. A1.1-1, although the E. D. C. can be equally applied to longitudinally non-uniform structures by considering z as a parameter. The three-dimensional scalar wave equation that should be used in this case is

$$\nabla^2 F(x,y,z) + k_o^2 \varepsilon(x,y,z) F(x,y,z) = 0 \qquad \text{(A1.1 - 1)}$$

where for the present case $\varepsilon(x, y, z) = \varepsilon(x, y) = \varepsilon_q(x)$ $(q = 1, 2, 3)$ is the stepwise constant (along the y-axis) (complex) dielectric distribution of the waveguide. Assume an ansatz of the type

$$F(x,y,z) = g(x;y) f(x) e^{-i\beta z} \qquad \text{(A1.1 - 2)}$$

where g(x;y) and f(x) are the vertical (parametrised in y) and horizontal mode distributions respectively, with g(x;y) slowly varying along x. Substitute (A1.1 - 2) into (A1.1 - 1), and obtain an equation that can be approximated by

$$\frac{1}{g(x;y)} \partial_y^2 g(x;y) + \frac{1}{f(x)} \partial_x^2 f(x) + \left(k_o^2 \varepsilon(x;y) - \beta^2 \right) = 0 \qquad \text{(A1.1 - 3)}$$

In (A1.1 - 3) the terms $f(x)\partial_x^2 g(x;y)$ and $f(x)\partial_x g(x;y)$ have been neglected. Thus, first solve

$$\partial_y^2 g(x;y) + \left[k_o^2 \varepsilon(x,y) - \beta_{eff}^2(x)\right] g(x;y) = 0 \qquad \text{(A1.1 - 4)}$$

to derive the vertical modes g(x;y) and the corresponding effective refractive indices in each region in which ε(x, y) is constant. The resulting $\beta_{eff}(x)$ are then used to solve the remaining equation for the transverse modes:

$$\partial_x^2 f(x) + \left[k_o^2 \varepsilon_{eff}(x) - \beta^2\right] f(x) = 0 \qquad \text{(A1.1 - 5)}$$

where $k_o^2 \varepsilon_{eff}(x) = \beta_{eff}^2(x)$. Hence, $\varepsilon_{eff}(x)$ is the dielectric distribution that is used to define the equivalent 'effective' two-dimensional structure.

With reference to Fig. A1.1-1, the three regions are defined by $x < x_1$, $x_1 < x < x_2$, $x > x_3$ and the dielectric distribution is stepwise constant in each region, i.e., $\varepsilon_1(y)$, $\varepsilon_2(y)$, $\varepsilon_3(y)$. Hence, equation (A1.1 - 4) becomes

$$\partial_y^2 g_q(y) + \left[k_o^2 \varepsilon_i(y) - \beta_{eff,q}^2\right] g_q(y) = 0 \qquad \text{(A1.1 - 6)}$$

with q = 1, 2, 3. From the solution of (A1.1 - 6) obtain the effective refractive indices of each region $\beta_{eff,q} = k_o n_{eff,q}$. The following equation is thus to be solved to find the transverse modes:

$$\partial_x^2 f(x) + \left[k_o^2 \varepsilon_{eff,q} - \beta^2\right] f(x) = 0 \qquad \text{(A1.1 - 7)}$$

where $\varepsilon_{eff,q} = n^2_{eff,q}$ is the effective dielectric distribution of the equivalent two-dimensional structure shown in Fig. A1.1-2. The use of equation (A1.1 - 7) implies that the structures of Fig. A.1.1-2 is considered as the two-dimensional equivalent of the structure shown in Fig. A1.1-1.

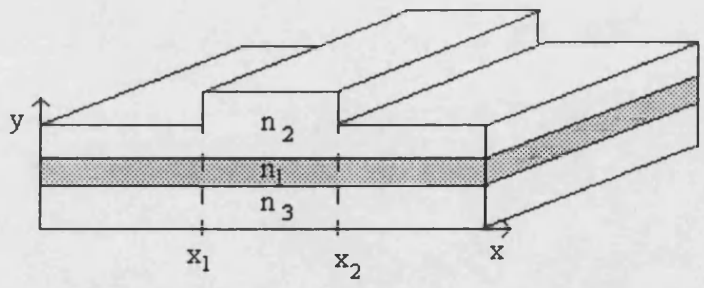


Fig. A1.1-1: Typical semiconductor device etched from multilayer material. In this specific example the variation of the thickness of the layers produces the refractive index step.

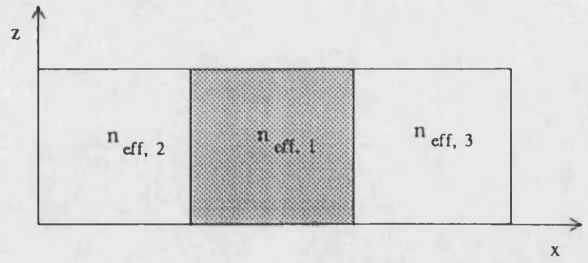


Fig. A1.1-2: Two-dimensional equivalent structure of the one shown in Fig. A1.1-1 obtained using the E. D. C. method.

Appendix 1.2

Plane wave decomposition and the paraxial wave equation

In a homogeneous medium with refractive index n_h (or $\varepsilon_h = n_h^2$), a general field function $F(x,z)$ may be written in the form of a Fourier Integral:

$$F(x,z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} A(k_x) e^{-i(k_x x + k_z z)} dk_x \qquad \text{(A1.2 - 1)}$$

where $A(k_x)$ is the angular spectrum distribution, and k_x and k_z the transverse and longitudinal components of the field propagation vector. Substituting (A1.2 -1) into the wave equation (1.4.4) it follows that the two components of the wave vector are related by the condition $k_x^2 + k_z^2 = k_o^2 \varepsilon_h$, and hence

$$k_z = \pm\sqrt{k_o^2 \varepsilon_h - k_x^2} \qquad \text{(A1.2 - 2)}$$

In the range for which the paraxial approximation is valid, the square root in equation (A1.2 - 2) is real. The paraxial wave equation, in fact, describes waves that are characterised by a wave vector almost parallel to the axis of propagation (the z-axis, in this case). Thus, the individual plane wave components in (A1.2 - 1) may be viewed as propagating at small angles to the longitudinal axis (z), Fig. A1.2-1, [16], [17].

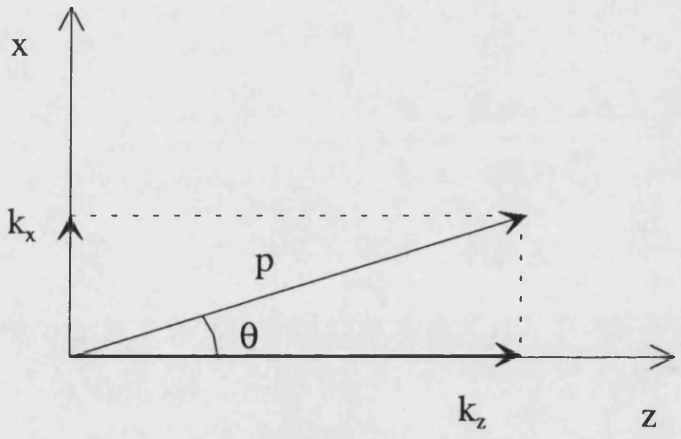


Fig. A1.2-1: Decomposition on the wave vector $\underline{p}$ ($p = k_o\varepsilon_h$) along the transverse (x) and longitudinal (z) axes.

From (A1.2 - 1), the initial field distribution is

$$F(x,z=0)=F_o(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}A(k_x)e^{-ik_x x}dk_x \qquad (A1.2-3)$$

By recognising that equation (A1.2 - 3) is a Fourier transform, by inverse transformation obtain the angular spectrum:

$$A(k_x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}F_o(x)e^{+ixk_x}dx \qquad (A1.2-4)$$

In the particular case of a Gaussian initial field distribution

$$F(x,z=0)=F_o(x)=E_o e^{\frac{-x^2}{2w_o^2}} \qquad (A1.2-5)$$

where E_o and w_o are constants, the angular spectrum distribution is

$$A(k_x)=E_o w_o e^{-\frac{1}{2}w_o^2 k_x^2} \qquad (A1.2-6)$$

which is a Gaussian function of width $\frac{\sqrt{2}}{w_o}$.

Using (A1.2 - 1), the field distribution at any longitudinal position z is

$$F(x,z)=\frac{1}{\sqrt{2\pi}}E_o w_o\int_{-\infty}^{+\infty}e^{-\frac{1}{2}w_o^2k_x^2}e^{-ik_x x}e^{-iz\sqrt{(k_o^2\varepsilon_h-k_x^2)}}dk_x \qquad (A1.2-7)$$

Noting that the angular spectrum (A1.2 - 6) tends to zero rapidly outside a finite interval, it is possible to use the paraxial approximation, i.e., $k_x << k_o\varepsilon_h$. Hence from (A1.2 - 2), k_z becomes

$$k_z=\left(k_o^2\varepsilon_h-k_x^2\right)^{\frac{1}{2}}\approx k_o n_h-\frac{k_x^2}{2k_o n_h} \qquad (A1.2-8)$$

Making use of (A1.2 - 8) in equation (A1.2 - 7) the field at $z > 0$ is

$$F(x,z)=\frac{E_o w_o}{\sqrt{w_o^2-i\frac{z}{k_o n_h}}}\exp\left(-\frac{x^2}{2\left(w_o^2-\frac{iz}{k_o n_h}\right)}\right)\exp(-izk_o n_h) \qquad (A1.2-9)$$

With a few algebraic steps (A1.2 - 9) can be reduced to the expression for the (fundamental) Gaussian Beam [refer to Chapter 2]. This result is valid also for higher order Gaussian Beams. In summary, the Gaussian Beams are usually referred to as free-space modes because the diffracted field originated from a Gaussian Beam is still a Gaussian Beam of the same order, but characterised by different size and phase, [17].

It can also be demonstrated that the paraxial wave equation in a homogeneous medium and the Huygens-Fresnel-Kirchhoff integral represent the same mathematical (and physical) approximation, [17]. Free-space propagation may be analysed using the paraxial wave equation, or, alternatively, the integral equation based on Huygens' principle in the Fresnel approximation. Huygens' principle states that given the field distribution $F_o(x, z)$ incident on a closed surface S_o, each point on this surface may be considered as the source of a uniform spherical wave (Huygens' wavelet). The total field at any point in space can therefore be calculated by summing the fields of the wavelets originating from all the point sources on S_o. Huygens' principle can be expressed in mathematical form (in a Cartesian two-dimensional representation), as

$$F(x,z) \approx \sqrt{\frac{i}{\lambda(z-z_o)}} e^{-ip(z-z_o)} \int F_o(x_o,z_o) e^{-ip\frac{(x-x_o)^2}{2(z-z_o)}} dx_o \qquad \text{(A1.2 - 10)}$$

where λ is the wavelength in the homogeneous medium. Hence, given the initial field distribution $F_o(x_o, z_o)$, the field $F(x, z)$ at any (x, z) is obtained by solving the integral in equation (A1.2 - 10). The solution is accurate within the paraxial approximation.

Appendix 1.3

Modes of parallel plate metal waveguides

1. TM modes

Consider a parallel plate metal guide of the type shown in Fig. 1.9.1. The characteristic equation for TM modes describes the longitudinal component of the electric field, $E_z(x)$, the other finite field components being E_x and H_y, [1], [3]. Assuming a mode solution $E_z(x, z) = e_z(x)e^{-i\beta z}$, (1.6.1), the scalar wave equation (1.4.4) becomes

$$\frac{d^2e_z}{dx^2} + k_c^2 e_z = 0 \qquad \text{(A1.3 - 1)}$$

with $k^2_{c(m)} = k_o^2\varepsilon - \beta_m^2$ the characteristic eigenvalue which determines the mode propagation constant, β_m. The boundary conditions associated with (A1.3 - 1) at the perfectly conducting boundaries are expressed by equation (1.2.5). Hence, $e_z = 0$ at $x = -a$ and $x = +a$, which imposes

$$e_z(x) = A\cos(k_{c(m)}x) \qquad \text{(A1.3 - 2)}$$

and also that the characteristic eigenvalues of equation (A1.3 - 1) take on discrete values determined by

$$k_{c(m)} = \frac{m\pi}{2a} \qquad \text{(A1.3 - 3)}$$

with $m = 1, 3, 5, \ldots$.

The propagation constant, β_m, for the m-th order TM mode is

$$\beta_m = \sqrt{k_o^2\varepsilon - k_{c(m)}^2} = \sqrt{\frac{\omega^2}{c^2}\varepsilon - \left(\frac{m\pi}{2a}\right)^2} \qquad \text{(A1.3 - 4)}$$

The other field components of TM modes derive from Maxwell's equations, i.e.,

$$h_y(x) = -i\frac{\omega\varepsilon_o\varepsilon}{k_o^2\varepsilon - \beta_m^2}\partial_x e_z \qquad \text{(A1.3 - 5)}$$

$$e_x(x) = \frac{i}{\omega\varepsilon\varepsilon_o}\partial_z H_y = \frac{\beta_m}{\omega\varepsilon\varepsilon_o}h_y(x) \qquad \text{(A1.3 - 6)}$$

where $H_y(x,z) = h_y(x)e^{-i\beta_m z}$, and $m = 1, 3, 5, \ldots$.

Imposing $\beta_m = 0$ in (A1.3 - 4) obtain the cut-off frequency:

$$\omega_{c(m)} = \frac{m\pi}{2a}\frac{c}{\sqrt{\varepsilon}} \qquad \text{(A1.3 - 7)}$$

which determines the frequency at which the m-th mode ceases to propagate ($\beta_m = 0$), [9]. For frequencies below cut-off β_m becomes purely imaginary and hence the mode attenuates rather than propagates (in which case β_m is called attenuation constant), [9]. In general, waveguides are operated in a low frequency range where only one propagating mode exists in the waveguide in order to avoid problems associated with modal dispersion typical of multimode operation, [1].

In many cases, e.g., reflection problems, it is useful to consider the wave impedance parameter along the direction of propagation, which is defined as the ratio of the transverse electric field to the transverse magnetic field for the individual mode, [3]. For the m-th order TM mode this parameter is

$$Z_{(m)}^{TM} = \frac{e_x}{h_y} = \frac{\beta_m}{\omega\varepsilon\varepsilon_o} = \sqrt{\frac{\mu_o}{\varepsilon_o\varepsilon}}\sqrt{1-\left(\frac{\omega_{c(m)}}{\omega}\right)^2} \qquad \text{(A1.3 - 8)}$$

From (A1.3 - 8) it is seen that the characteristic impedance $Z_{(m)}^{TM}$ is independent of the transverse variable (x), and that for frequencies below cut-off it becomes purely imaginary, which corresponds to a situation of no power flowing along the direction of propagation. The wave impedance parameter is generally used to express the transmitted power of an individual mode in terms of the field transverse components only. In fact, the transverse component (A1.3 - 6) of the electric field of the m-th mode may be written as

$$e_x(x) = Z_{(m)}^{TM}\, h_y(x) \qquad \text{(A1.3 - 9)}$$

and hence the power carried by this TM mode is

$$P_{(m)} = \int_{-a}^{+a}\frac{\beta_m}{2\varepsilon_o\varepsilon}\left|h_y(x)\right|^2 dx = \tfrac{1}{2}Z_{(m)}^{TM}\int_{-a}^{+a}\left|h_y(x)\right|^2 dx \qquad \text{(A1.3 - 10)}$$

2. TE modes

TE modes are characterised by the three finite field components H_x, E_y, H_z. The scalar wave equation is now solved for the longitudinal component of the magnetic field, H_z, assuming a mode of the type (1.6.1). Following a derivation similar to that

used for TM modes, it is found that the z-component of the magnetic field of TE modes is

$$h_z(x) = B\sin(k_{c(m)}x) \qquad \text{(A1.3 - 11)}$$

with $k_{c(m)}$ given by

$$k_{c(m)} = \frac{m\pi}{2a} \qquad \text{(A1.3 - 12)}$$

where m = 1, 3, 5, ... The other field components of TE modes derive from Maxwell's equations and are

$$h_x(x) = -\frac{i}{\omega\mu_o}\partial_z E_y = -\frac{\beta_m}{\omega\mu_o}e_y(x) \qquad \text{(A1.3 - 13)}$$

$$e_y(x) = \frac{i\omega\mu_o}{k_o^2\varepsilon - \beta_m^2}\partial_x h_z(x) \qquad \text{(A1.3 - 14)}$$

where $E_y(x,z) = e_y(x)e^{-i\beta_m z}$.

The expression for the cut-off frequency, $\omega_{c(m)}$, for the attenuation constant (α_m) below cut-off and for the propagation constant (β_m) above cut-off are the same as those obtained for TM modes, i.e., equations (A1.3 - 7) and (A1.3 - 4) respectively.

The characteristic impedance for TE modes is defined by

$$Z_{(m)}^{TE} = -\frac{e_y}{h_x} = \frac{\omega\mu_o}{\beta_m} = \sqrt{\frac{\mu_o}{\varepsilon_o\varepsilon}}\frac{1}{\sqrt{1-\left(\frac{\omega_{c(m)}}{\omega}\right)^2}} \qquad \text{(A1.3 - 15)}$$

which, as for TM modes, is independent of the transverse variable (x) and becomes purely imaginary for frequencies below cut-off. The power carried by TE modes is

$$P_{(m)} = \int_{-a}^{+a}\frac{\beta_m}{2\omega\mu_o}|e_y(x)|^2\,dx = \frac{1}{2Z_{(m)}^{TE}}\int_{-a}^{+a}|e_y(x)|^2\,dx \qquad \text{(A1.3 - 16)}$$

Appendix 1.4

Functional form of TE and TM modes of the three-layer symmetric slab dielectric waveguide - scalar analysis

For (longitudinally uniform) slab waveguide structures, Fig. 1.10.1, in which the refractive index profile is a function of only the transverse variable (x), two separate sets of equations can be derived from Maxwell's equations, one of which determines the $(TE)_x$ while the other corresponds to the $(TM)_x$ modes, [4]. The 'generating' field components for the TE and the TM modes may, e.g., be chosen to be, respectively,

$$E_y = e_y(x)e^{-i\beta_{(e)}z} \qquad \text{(A1.4 - 1)}$$

$$H_y = h_y(x)e^{-i\beta_{(h)}z} \qquad \text{(A1.4 - 2)}$$

with $e_y(x)$ and $h_y(x)$ the transverse distributions of the electric and magnetic field components and $\beta_{(e)}$ and $\beta_{(h)}$ the corresponding propagation constants. They satisfy, respectively, the scalar wave equations, [24],

$$\nabla^2 E_y + k_o^2 \varepsilon(x) E_y = 0 \qquad \text{(A1.4 - 3)}$$

$$\nabla^2 H_y + \frac{1}{\varepsilon}\partial_x \varepsilon \partial_x H_y + k_o^2 \varepsilon(x) H_y = 0 \qquad \text{(A1.4 - 4)}$$

Substituting (A1.4 - 1) and (A1.4 - 2) into equations (A1.4 - 3) and (A1.4 - 4), respectively, yield the equation describing the transverse distributions of the y-field components as follows:

$$\frac{d^2 e_y}{dx^2} + \left(k_o^2 \varepsilon(x) - \beta_{(e)}^2\right) e_y = 0 \qquad \text{(A1.4 - 5)}$$

$$\frac{d^2 h_y}{dx^2} + \frac{1}{\varepsilon(x)}\frac{d\varepsilon}{dx}\frac{dh_y}{dx} + \left[k_o^2 \varepsilon(x) - \beta_{(h)}^2\right] h_y = 0 \qquad \text{(A1.4 - 6)}$$

Equation (A1.4 - 5) is used to solve for the TE modes, while equation (A1.4 - 6) for the TM modes.

In particular, for cases in which the $\underline{\nabla}\varepsilon$ term is neglected in (A1.4 - 4), the scalar wave equation (1.4.4) provides approximate (degenerate) solutions for the two polarisations that are accurate for most practical situations, i.e., the propagation constants (and related field profiles) of TE and TM modes become identical. In many weak dielectric step optical waveguides, e.g., the neglect of $\underline{\nabla}\varepsilon$ is justified, [4].

In the case of slab waveguides with (multi-) step index profile, the modes are obtained by solving the scalar wave equation in each region in which the refractive index is piece-wise constant (since the $\underline{\nabla}\varepsilon$ term in each such regions is identically zero). The distinction between TE and TM modes arises when the matching condition for both the electric and the magnetic field at the interface between regions with different values of the dielectric constant are taken into account.

Once the equations for E_y and H_y are solved, with the appropriate boundary conditions, [Section 2], the other field components are obtained using Maxwell's equations (1.1.1) - (1.1.4); hence, the equations defining TE and TM modes result:

$$\text{TE modes} \quad \begin{cases} H_x = -\dfrac{\beta_{(e)}}{\omega\mu_o} E_y \\ H_z = \dfrac{i}{\omega\mu_o} \partial_x E_y \end{cases} \qquad \text{(A1.4 - 7)}$$

and

$$\text{TM modes} \quad \begin{cases} E_x = \dfrac{\beta_{(h)}}{\omega\varepsilon_o n^2(x)} H_y \\ E_z = -\dfrac{i}{\omega\varepsilon_o n^2(x)} \partial_x H_y \end{cases} \qquad \text{(A1.4 - 8)}$$

Thus, for TE modes, the electric field has no longitudinal component, and the electromagnetic field is characterised by the three components H_x, E_y, H_z. Analogously, for TM modes, the magnetic field has no longitudinal component, and the electromagnetic field has only three components, i.e., E_x, H_y, E_z.

It is worth noting that, when the dielectric distribution is symmetric, i.e., $\varepsilon(-x) = \varepsilon(x)$, the TE and the TM modes are either symmetric or antisymmetric in x.

1. TE modes (H_x, E_y, H_z)

i) Bound modes

$$e_y(x) = \begin{cases} A\begin{pmatrix} \sin(k_{x1}x) \\ \cos(k_{x1}x) \end{pmatrix} & |x| \le a \\ A\begin{pmatrix} \dfrac{x}{|x|}\sin(k_{x1}a) \\ \cos(k_{x1}a) \end{pmatrix} e^{-k_{x2}(|x|-a)} & |x| \ge a \end{cases} \qquad \text{(A1.4 - 9)}$$

with $k_{x1}^2 = k_o^2\varepsilon_1 - \beta_m^2$ and $k_{x2}^2 = \beta_m^2 - k_o^2\varepsilon_2$. The upper functional form refers to antisymmetric modes, while the lower one to symmetric modes. The complete expression for TE bound modes is $E_y(x) = e_y(x)e^{-i\beta_m z}$, m = 0, 1, 2, ...

Condition for bound modes: $k_{x1}^2, k_{x2}^2 > 0$

ii) Radiation modes

$$e_y(x) = \begin{cases} C\begin{pmatrix}\sin(\sigma x)\\ \cos(\sigma x)\end{pmatrix} & |x| \le a \\ \begin{pmatrix}\frac{x}{|x|}\\ +\end{pmatrix}\left(De^{-ipx} + Fe^{ipx}\right) & |x| \ge a \end{cases} \qquad \text{(A1.4 - 10)}$$

with $\sigma^2 = k_o^2\varepsilon_1 - \beta^2$ and $\rho^2 = k_o^2\varepsilon_2 - \beta^2$. The upper functional form refers to antisymmetric modes, the lower one to symmetric modes. The complete expression for TE radiation modes is $E_y(x) = e_y(x)e^{-i\beta z}$.

Condition for radiation modes: $\rho^2 > 0$.

The other field components of TE modes derive from Maxwell's equations:

$$h_x(x) = -\frac{i}{\omega\mu_o}\frac{\partial E_y}{\partial z} \qquad \text{(A1.4 - 11)}$$

$$h_z(x) = \frac{i}{\omega\mu_o}\frac{\partial e_y}{\partial x} \qquad \text{(A1.4 - 12)}$$

The power carried by TE (guided) modes is

$$P_{(m)} = \int_{-\infty}^{+\infty} \frac{\beta_m}{2\omega\mu_o}\left|e_y(x)\right|^2 dx \qquad \text{(A1.4 - 13)}$$

2. TM modes (E_x, H_y, E_z)

The functional forms for TM modes are as described above but replacing $e_y(x)$ with $h_y(x)$; the other field components for TM modes are

$$e_x(x) = \frac{i}{\omega\varepsilon_o\varepsilon}\frac{\partial H_y}{\partial z} \qquad \text{(A1.4 - 14)}$$

$$e_z(x) = -\frac{i}{\omega\varepsilon_o\varepsilon}\frac{\partial h_y}{\partial x} \qquad \text{(A1.4 - 15)}$$

where $H_y(x) = h_y(x)\exp(-i\beta_m z)$. Note that TM modal fields in three-layer slab dielectric waveguides are discontinuous because the dielectric distribution $\varepsilon(x)$ is discontinuous (e.g., in a three-layer slab waveguide the dielectric distribution is stepwise constant).

The power carrier by a TM (guided) mode is

$$P_{(m)} = \int_{-\infty}^{+\infty} \frac{\beta_m}{2\varepsilon_o \varepsilon(x)} \left| h_y(x) \right|^2 dx \qquad \text{(A1.4 - 16)}$$

References

[1] R. E. Collin, *Field theory of guided waves*, IEEE Press (1990)

[2] R. März, *Integrated optics*, Artech House (1995)

[3] S. Ramo, J. R. Winnery, T. Van Duzer, *Fields and waves in communication electronics*, John Wiley and Sons (1990)

[4] D. Marcuse, *Light transmission optics*, Van Nostrand Reinholds (1982)

[5] T. Rozzi, M. Mongiardo, *Open dielectric waveguides*, IEE Electromagnetic wave series (1997)

[6] A. K. Ghatak, K. Thyagarajan, *Optical electronics*, Cambridge University Press (1989)

[7] P. Diament, *Wave transmission and fiber optics*, Macmillan Publishing Company (1990)

[8] G. H. Owyang, *Foundations of optical waveguides*, Edward Arnold Publishers (1981)

[9] S. F. Mahmoud, *Electromagnetic waveguides: theory and applications*, Peter Peregrinus (1991)

[10] D. Marcuse, *Solution of the wave equation for general dielectric waveguides by the Galerkin method*, IEEE Journal of Quantum Electronics, 28, (1992), pp. 459-465

[11] A. Weisshaar, J. Li, R. Gallawa, I. C. Goyal, *Vector and quasi vector solutions for optical waveguide modes using efficient Galerkin's method with Hermite-Gauss basis functions*, Journal of Lightwave Technology, 13 (1995), pp. 1795-1800

[12] H. C. Chen, *Theory of electromagnetic waves - A coordinate-free approach*, McGraw-Hill Book Company (1983)

[13] L. B. Felsen, N. Marcuvitz, *Radiation and scattering of waves*, Prentice-Hall (1973)

[14] A. Sommerfeld, *Partial differential equations in Physics. Lectures on theoretical Physics - vol VI*, Academic Press (1964)

[15] D. C. Stinson, *Intermediate mathematics of electromagnetics*, Prentice-Hall Electrical Engineering Series (1976)

[16] H. E. Haus, *Waves and fields in optoelectronics*, Prentice-Hall series in solid state physical electronics (1984)

[17] A. E. Siegman, *Lasers*, University Science Books (1986)

[18] F. Sporleder, H. G. Unger, *Waveguide tapers transitions and couplers*, Peter Peregrinus (1979)

[19] V. V. Shevchenko, *Continuous transitions in open waveguides*, The Golem Press (1971)

[20] D. Marcuse, *Theory of dielectric optical waveguides*, Academic Press (1974)

[21] H. Kressler, J. K. Butler, *Semiconductor lasers and heterojunction LEDs*, Academic Press (1977)

[22] G. P. Agrawal, N. K. Dutta, *Long-wavelength semiconductor lasers*, Van Nostrand Reinhold (1986)

[23] T. Tamir ed., *Guided-wave optoelectronics*, Springer-Verlag (1990)

[24] A. K. Ghatak, M. S. Sodha, *Inhomogeneous optical waveguides*, Plenum Press (1977)

Chapter 2

The Hermite-Gauss functional form

In this chapter the relevant properties of the two distinct sets of the Hermite-Gauss Beams (HGB) and the Hermite-Gauss Eigenfunctions (HGE) are presented. The purpose of this review is to highlight the differences between the two sets of functions, and to introduce the HGEs which will form the basis set for the method developed in this thesis for the analysis of field propagation in weakly guiding structures. The following discussions are with reference to the two-dimensional space described by the Cartesian co-ordinates x and z.

Although both the HGB and the HGE function sets are complete and have nearly the same nomenclature, which denotes the functional form, they are, in fact, characterised by different properties that can be related to the equations from which they originate. The HGBs are the solutions of the paraxial wave equation in free-space (unbounded, homogeneous medium). Hence, they are the solutions of a first order partial differential equation and constitute only an approximate solution of the complete wave equation. Specifically, the HGBs are characterised by a Hermite-Gaussian profile whose size changes with distance along the direction of propagation. This means that each HGB function depends on both x and z to represent diffraction effects.

The Hermite-Gauss Eigenfunctions are of a different nature. They are the eigensolutions of the wave equation in a quadratic medium[1] which is a second order differential equation of the Sturm-Liouville type. As with the HGBs, the HGEs are characterised by a Hermite-Gaussian profile, but now the profile is unchanged and the phase front remains flat[2] during propagation in the quadratic medium because here diffraction effects are counteracted by the focusing effect of the quadratically varying

[1] A quadratic medium is an inhomogeneous medium with quadratically varying (parabolic) dielectric profile.

[2] For a mode (an individual HGE is a mode of the quadratic medium, as discussed below), the (complex) modal field may be written in the form $F_m(x,z) = |f_m(x)|\exp(i\theta_m(z))$, where $f_m(x)$ is the mode transverse profile, and $\theta_m(z)$ is the phase front which is independent of x and such that $\theta_m(z) = \beta_m z$. Consequently, at each longitudinal position z = const the phase front is $\theta_m(z)$ = const which is usually referred to as 'flat' (or 'plane') phase front.

dielectric profile. The HGEs are also known as harmonic functions (eigensolutions of the harmonic oscillator of quantum mechanics, [1]).

1. Weak diffraction in homogeneous media: the Hermite-Gauss Beams

Consider a homogeneous medium characterised by the constant dielectric distribution $\varepsilon(x,z)=\bar{\varepsilon}$ and, for convenience, define $p^2 = k_o^2\bar{\varepsilon}$. By substitution of the ansatz $F(x, z) = f(x, z)e^{-ipz}$ into the paraxial wave equation (1.7.6) obtain

$$\partial_z f(x,z) = -\frac{i}{2p}\partial_x^2 f(x,z) \qquad (2.1.1)$$

It can be shown, [Appendix 1.2], that the solutions of (2.1.1) are the Hermite-Gauss Beams, [2], [3], [4]. The characteristic functional form for a HGB is

$$f_n(x,z) = \left(\frac{1}{\sqrt{\pi}\,2^n n!}\right)^{\frac{1}{2}}\left(\frac{w_o}{w(z)}\right)^{\frac{1}{2}} H_n\left(\frac{x}{w(z)}\right) e^{-\frac{x^2}{2w^2(z)}} e^{-ip\frac{x^2}{2R(z)}} e^{-i\varphi_n(z)} \qquad (2.1.2)$$

where $H_n(s)e^{-\frac{1}{2}s^2}$ is the Hermite-Gauss function of integer order, n;

$w(z) = w_o\sqrt{1+\left(\frac{z}{z_o}\right)^2}$ is the half beam-width whose minimum value (w_o) is referred to as the beam waist - note that w(z) is independent of the mode number n;

$R(z) = z+\frac{z_o^2}{z}$ is the radius of curvature of the phase front which also is independent of the mode number;

$\varphi_n(z) = \frac{n+1}{2}\operatorname{arctg}\left(\frac{z}{z_o}\right)$ is a term contributing to the phase of the wave which does depend on the mode order n;

$p = \frac{2\pi}{\lambda}$, with λ the wavelength in the medium;

$z_o = \frac{\pi w_o^2}{\lambda}$ is the Rayleigh range.

The smaller the spotsize w_o at the beam waist, the smaller the Rayleigh range, and hence the greater the rate of growth with z of the spot size from the waist, [3]. It is possible to define a divergence angle for the Gaussian Beam as

$\theta \approx \frac{w(z)}{z} \approx \frac{w_o}{z_o} = \frac{\lambda}{\pi w_o}$ (for $z >> z_o$) which is of the same order of the divergence angle associated with the diffraction of a plane wave by an aperture of diameter ~w_o located at the beam waist, [3].

To visualise such two dimensional functions, note that for any z = const., the field amplitude has a Hermite-Gaussian distribution along x, with a waist size $w = w(z)$ corresponding to that longitudinal position. Further, at any specified longitudinal position, z, the phase front has a defined curvature [note in equation (2.1.2) the additional phase terms which depend on both z and the mode number]. From equation (2.1.2) it can be seen that the phase front is flat only at the beam waist (when z = 0, $w(z) = w_o$) while at large distances from the beam waist plane it is curved (since $R(z) \approx z$, for $z >> z_o$) and centred at the beam waist.

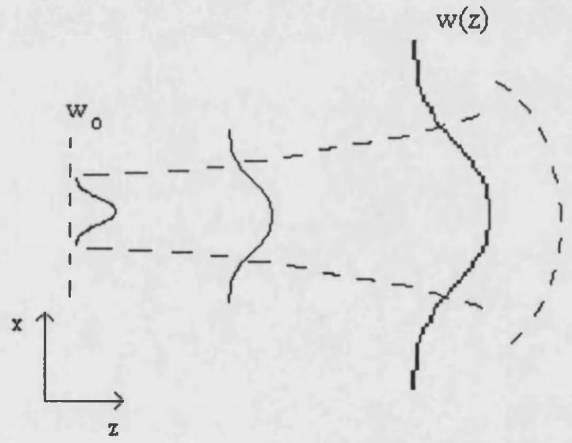


Fig. 2.1.1: Diffraction of the fundamental Gaussian Beam (propagation is assumed to be only in the +z-direction). The broken curves show the phase front which is flat at the beam waist ($w = w_o$) and curved at $z > 0$, and also the beam waist increasing with z.

The fact that both the width of the beam and its radius of curvature are independent of the mode number is relevant for practical problems of mode matching, e.g., in the coupling of single mode fibres to laser, [2].

At any z = const. it is possible to show that the HGBs (2.1.2) satisfy the following differential equation of the Sturm-Liouville type

$$\left[\frac{d^2}{dx^2} - x^2\right] g_n(x) = -(2n+1) g_n(x) \qquad (2.1.3)$$

where $g_n(x) = f_n(x, z = \text{const.})$ of equation (2.1.2). Using (2.1.3) it can be demonstrated that (at any z = const.) the HGB set is complete in the vector space of square integrable functions[1] [5] since the operator $\left[\frac{d^2}{dx^2} - x^2\right]$ in equation (2.1.3) is hermitian[2].

The HGBs have been widely used in the analysis of solid state and gas laser beams, [6] as well as in the theory of open resonators, [4], [7]. Also, in optics, the HGBs represent an invaluable means of analysis since the laws of propagation for the HGBs in a variety of optical systems (e.g., homogeneous media, interfaces between two homogeneous media, lenses, and apertures) can be synthesised within the ABCD matrix method, [4], [8]. The ABCD matrix method is a fast and convenient formalism for describing the transformation of HGBs travelling through a combination of optical systems which can be represented by a succession of lenses and free-space elements (under the geometrical optics approximation). These properties make the HGBs ideal for the analysis of quasi-optical propagation; thus they find widespread application in the design, [9], and also the analysis of far-field characteristics, [10], [11], of microwave antennas.

The HGBs may also be derived from Huygens' diffraction integral instead of via the paraxial wave equation [Appendix 1.2], since the two analyses are equivalent in homogeneous media, [4]. The diffraction integral approach has also been applied to the analysis of open resonators [3], [4], [6]. In this particular case, the modes are field distributions that repeat themselves at each round-trip. Most lasers operate with apertures which are large compared to the (optical) wavelength so that the output beams may be accurately described by Hermite-Gauss Beams. It is found that by

---

[1] Square integrable functions, f(x), over an interval (a,b) have the property that the integral $\int_a^b |f(x)|^2 dx$ is finite. The set of square integrable functions is a vectorial space with inner product defined by $(f,g) = \int_a^b f^*(x) g(x) dx$ (where * denotes the complex conjugate).

[2] Complex self-adjoint operators are usually called hermitian operators, and they are defined by $A = A^\dagger$, with $A^\dagger$ the hermitian operator († is the operation of Hermitian conjugation, i.e., complex conjugation plus transposition of variables). The most relevant properties of such operators are that all eigenvalues are real, and that eigenfunctions corresponding to different eigenvalues are orthogonal, [5]. The class of Sturm-Liouville operators are one example of hermitian operators.

placing mirrors of proper curvature to match the curved phase front of the HGB at a specified longitudinal position it is possible to produce a variety of laser resonator configurations, [2], [3].

2. Modes of quadratic media: the Hermite-Gauss Eigenfunctions

For the following discussion consider a non-homogeneous medium with a continuously varying dielectric distribution that has a quadratic (parabolic) transverse variation of the form $\varepsilon(x) = b - ax^2$, with a and b positive real. Such a medium is generally referred to as a quadratic medium. In this case it can be shown that the refractive index profile is also parabolic. By writing $n(x) = n_o - n_1 x^2$ with $n_1 << 1$, the dielectric distribution becomes $\varepsilon(x) \approx n_o^2 - 2n_o n_1 x^2$, in which case $a = 2n_o n_1$ and $b = n_o^2$. (Note that although a quadratic dielectric profile is not physically realisable in the full range $0 < |x| < \infty$, such a distribution is often considered for mathematical convenience, [12]).

A quadratic medium can be viewed as a succession of an infinite number of thin lenses whose distance from one another has been reduced to zero. This produces a continuously guiding medium; for example, graded index optical glass fibres have a refractive index distribution of the kind described above, [2]. Strictly speaking, for media with continuously varying refractive index profile the Helmholtz equation (1.4.4) is not an exact equivalent of Maxwell's equations, [13], [12]. However, considering a refractive index distribution that varies slowly with x, the scalar wave equation approach considered in this discussion is an acceptable approximation.

The scalar wave equation (1.4.4) for quadratic media takes the form

$$\left[\partial_x^2 + \partial_z^2 + k_o^2\left(b - ax^2\right)\right]F(x,z) = 0 \qquad (2.2.1)$$

For mode-type solutions of the general form

$$F(x,z) = F_m(x,z) = f_m(x)e^{-i\beta_m z} \qquad (2.2.2)$$

the eigenvalue equation is

$$\frac{d^2 f_m(x)}{dx^2} + k_{x,m}^2 f_m(x) = 0 \qquad (2.2.3)$$

where $k_{x,m}^2(x) = \left[k_o^2(b - ax^2) - \beta_m^2\right]$ determines the eigenvalue, β_m, of the m-th eigenmode, $f_m(x)$. To solve equation (2.2.1), the following boundary conditions are assumed: $|f_m| \rightarrow 0$ and $\frac{df_m(x)}{dx} \rightarrow 0$ for $|x| \rightarrow \infty$ (refer to Sommerfeld radiation condition, [14], [15]) which hold for bound mode field solutions, as discussed in Chapter 1.

The eigensolutions of (2.2.3) are the Hermite-Gauss functions

$$f_m(x) = C_m H_m\left(\tfrac{x}{w_o}\right) e^{\frac{-x^2}{2w_o^2}} \tag{2.2.4}$$

where m is a positive integer, $C_m = \left(\frac{1}{w_o 2^m m! \sqrt{\pi}}\right)^{\frac{1}{2}}$ is the normalisation constant, $H_m\left(\frac{x}{w_o}\right)$ is the Hermite polynomial of order m, and w_o the constant specified by $w_o = \sqrt{\frac{1}{k_o\sqrt{a}}}$, is often referred to as the width parameter.

The eigenvalue equation (2.2.3) can be solved only if the eigenvalues are discrete, i.e., by imposing the condition $\frac{k_o^2 b - \beta_m^2}{k_o\sqrt{a}} = 2m + 1$, [5]. Hence the modes of a square-law medium are of the type described in equation (2.2.2) with transverse profile given in equation (2.2.4), Fig. 2.2.1, and have corresponding propagation constant

$$\beta_m = k_o \sqrt{b - \frac{\sqrt{a}}{k_o}(2m+1)} \tag{2.2.5}$$

It is worth noting that the HGEs are purely real and hence, viewed in conjunction with equation (2.2.2), they are characterised by plane phase fronts at z = const.

The eigenvalue equation (2.2.3) can also be written in the form

$$\left[\frac{d^2}{ds^2} + (-s^2 + 2m + 1)\right] f_m(s) = 0 \tag{2.2.6}$$

where $s = \frac{x}{w_o}$; in this form the equation is recognised to be of the Sturm-Liouville type, as equation (2.1.3). The operator $\left[\frac{d^2}{ds^2} - s^2\right]$ in (2.2.6) is hermitian; the HGEs,

thus, form a complete orthonormal set in the functional space of square integrable functions, [2], [4], [16].

The orthogonality condition for the HGEs is

$$\forall m,n \qquad \int_{-\infty}^{+\infty} f_m(x) f_n^*(x) dx = \delta_{mn} \tag{2.2.7}$$

which follows from the properties of the operator in (2.2.6).

Because of the following properties

i) completeness in the space of square summable functions,

ii) satisfying the radiation boundary conditions required for electromagnetic fields

iii) forming a (complete) discrete set

the HGEs provide an ideal basis set for electromagnetic field analysis

For example, in the case of the modal solution, $F_m(x)$, of the eigenvalue equation (1.10.3) for slab waveguides, field analysis by means of function expansion methods is typically expressed in the form

$$F_m(x) = \sum_k a_k^{(m)} f_k(x) \tag{2.2.8}$$

with $a_k^{(m)}$ the expansion coefficients and $f_k(x)$ the basis functions. For (2.2.8) to be valid it is essential that the basis set be complete in the functional space of interest - this will be discussed in detail in Chapter 4. It is worth pointing out that the HGBs also form a complete and orthonormal set in the same functional space as the HGEs. In fact, the HGBs have often been used as a basis set typically for field problems in free-space or in paraxial optical systems, [4]. However, it is argued that, compared to the HGBs, the HGE set is far more manageable for the solution of field propagation in inhomogeneous media.

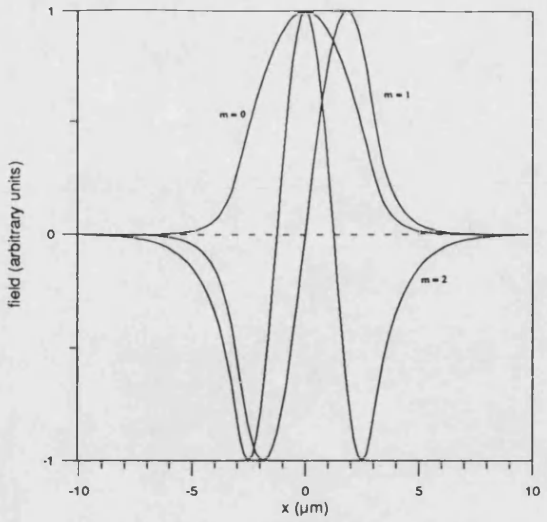


Fig. 2.2.1: The first three Hermite-Gauss Eigenfunctions.

3. Approximation of the three-layer slab dielectric waveguide modes by the HGEs - the variational method

The choice of the appropriate width parameter w_o can be of importance when using the HGEs (2.2.4) as the basis set for (total) field analyses of the type (2.2.8). In general no 'a priori' criteria exist for a convenient choice of w_o because of the several problem-dependent factors involved in such a choice; this aspect will be discussed in detail in Chapter 4. However, it is useful to present a short review of the variational method by means of which an optimal width parameter may be determined for the particular situation described below.

As evident from Fig. 2.2.1, the HGEs resemble the bound modes of the symmetric three-layer slab waveguide. This is the reason why they have often been used as an approximation of such modal profiles, [2]. In this case the variational method may be applied to calculate the best fitting HGE function for each corresponding waveguide bound mode, and hence the optimal width parameter (w_o) for that particular HGE is determined.

As an example, consider the fundamental mode of a symmetric three-layer slab waveguide which is to be approximated by a single Gaussian function, [2], [17], [18], [19]. Consider in the following discussion the waveguide parameters: n_1 and n_2 the refractive index in the core and cladding layers, respectively; core thickness = $2a_o$; λ_o

the free-space wavelength of the signal. The (approximated) fundamental mode is written as

$$\psi(x,z) = f(x)e^{-i\beta z} \tag{2.3.1}$$

where $f(x)=\sqrt{\frac{1}{w_o}\sqrt{\frac{2}{\pi}}}\exp\left(\frac{-x^2}{w_o^2}\right)$ is the HGE with the width parameter w_o and the propagation constant β to be determined. It can be shown, [2], that the scalar wave equation (1.4.4) is the Euler equation for the functional[1] $\Im = \iint\left[\left((\nabla\psi)\cdot(\nabla\psi^*)\right)-n^2k_o^2\psi\psi^*\right]dxdz$, whose stationary value is zero. Hence, every solution of the scalar wave equation (1.4.4) is also a stationary value for $\Im$, [2]. Strictly, this is only a necessary condition for $\Im$ to be an extremum, but it is assumed here that this condition is also sufficient, [2].

Assuming a normalised trial solution for $\Im$ of the type specified in equation (2.3.1), which is then automatically a solution of the scalar wave equation (1.4.4), and substituting it into the functional $\Im$, obtain

$$\beta^2 = \frac{\int_{-\infty}^{\infty}\left[n^2k_o^2f(x)f^*(x)-\partial_x f\partial_x f^*\right]dx}{\int_{-\infty}^{\infty}f(x)f^*(x)dx} \tag{2.3.2}$$

Equation (2.3.2) represents the variational expression for the mode propagation constant β. Since (2.3.1) is a stationary solution of the wave equation, β^2 should be insensitive to small variations of the solution, f(x), [18], [2], i.e.,

$$\frac{\partial\beta^2}{\partial w_o}=0 \tag{2.3.3}$$

Equation (2.3.3) is then used to determine the waist parameter, w_o, of the Gaussian that best approximates the mode field profile.

For the case of a symmetric three-layer step-index waveguide the (transcendental) equation for w_o is (in rectangular co-ordinates)

$$\sqrt{\frac{2}{\pi}}V^2e^{\frac{-2a_o}{w_o^2}}-\frac{a_o}{w_o}=0 \tag{2.3.4}$$

where $V=a_o\,k_o\sqrt{n_1^2-n_2^2}$ is the normalised frequency, [19].

[1] A number $\Im$ is a functional of f if its value depends on f, e.g., $\Im = |f|^2$.

The same variational technique has also been applied to analyse circular optical fibres, and in that case also accurate values for the propagation constant are obtained, [2], [4], [20].

The above analysis is easily extended to the higher order modes [19], but in general it is found that the values of the width parameter that optimise each HGE to the corresponding waveguide bound mode are different for modes of different order. Consequently, in general, it appears not to be practical to use the variational method for determining a single optimum width parameter for field expansions of the type (2.2.8).

4. Connection between the Hermite-Gauss Eigenfunctions and the Hermite-Gauss Beams

In this section an example of total field analysis in a quadratic medium is presented with the purpose of illustrating the relation between the HGEs (2.2.4) and the HGBs (2.1.2). An incident field $F(x, z=0) = F_o(x)$ in a quadratic medium is expanded in terms of the HGE modes, $f_m(x)$, supported by the medium as

$$F_o(x) = \sum_{m=0}^{\infty} a_m f_m(x) \tag{2.4.1}$$

where, applying (2.2.7),

$$a_m = \int_{-\infty}^{\infty} F_o(x) f_m(x) dx \tag{2.4.2}$$

Assume the medium to be specified by the refractive index profile $n(x) = n_o - n_1 x^2$, with $n_o > 0$ and $0 < n_1 << 1$. Since each individual HGE mode, $f_m(x)$, propagates with a propagation constant β_m specified by

$$\beta_m = \sqrt{k_o^2 n_o^2 - k_o \sqrt{2 n_o n_1}\,(2m+1)} \tag{2.4.3}$$

it is possible to write the (total) field at any z in the quadratic medium as

$$F(x,z) = \sum_m \left(\int_{-\infty}^{+\infty} F_o(x') f_m(x') dx' \right) f_m(x) e^{-i\beta_m z} \tag{2.4.4}$$

Using Mehler' s formula [21]

$$\sum_{n=0}^{\infty}\frac{c^n}{2^n n!}H_n(s)H_n(s')=\frac{\exp\left[\dfrac{2css'-c^2(s^2+s'^2)}{1-c^2}\right]}{\sqrt{1-c^2}} \qquad (2.4.5)$$

with $c=e^{i\sqrt{\frac{2n_1}{n_o}}z}$, $s=\dfrac{x}{w}$ and $w=\dfrac{1}{k_o\sqrt{2n_on_1}}$, equation (2.4.4) becomes

$$F(x,z)=\frac{e^{-ik_on_oz}e^{i\frac{\theta_c}{2}}}{w\sqrt{\pi(1-c^2)}}\int_{-\infty}^{+\infty}F_o(x')e^{i\frac{xx'}{w^2\sin\theta_c}}e^{-i\frac{x^2+x'^2}{2w^2\tan\theta_c}}dx' \qquad (2.4.6)$$

with $\theta_c=\sqrt{\dfrac{2n_1}{n_o}}z$.

In the limit of $n_1\to 0$, such that $\theta_c\to 0$, the focusing properties of the medium are negligible, i.e., the medium becomes homogeneous. This implies that the free-space propagation term, k_on_o, in equation (2.4.3) is the dominating term for β_m so that the individual HGEs are now approximately plane waves. In this case equation (2.4.6) can be reduced to

$$F(x,z)=\sqrt{\frac{-in_o}{\lambda_o z}}e^{-ik_on_oz}\int_{-\infty}^{+\infty}F_o(x')e^{-i\frac{k_on_o(x^2-x'^2)}{2z}}dx' \qquad (2.4.7)$$

Thus, as expected, in the case of a very weakly guiding medium ($n_1\approx 0$), the expression for determining the field propagating in a quadratic medium, equation (2.4.4), reduces to the Fresnel-Huygens' integral for calculating propagation in free-space. As already mentioned in Section 1, the HGBs are one set of eigensolutions of the latter equation. Thus, with the above derivation it has been shown that the HGEs lose the characteristics of guided eigensolutions and become diffracting HGBs during the transition from quadratic (guiding) medium to homogeneous (diffracting) medium.

Summary

In this chapter the two sets of functions, the Hermite-Gauss Beams (HGBs) and the Hermite-Gauss Eigenfunctions (HGEs), have been reviewed. It has been considered useful to recognise that they represent qualitatively different solutions of the wave equation. The HGBs are the solutions of the paraxial wave equation in a homogeneous medium; this is the reason why they are also known as free-space normal modes. The characteristic feature of the HGBs of diffracting during

propagation can be recognised in the longitudinally varying beam waist. On the other hand, the HGEs represent the modes of (longitudinally uniform) waveguides with quadratic dielectric distribution, and hence they are characterised by a constant beam waist.

Both sets of functions are complete in the functional space of interest, and hence both can be equally used as the basis set for total field expansion schemes. However, it is felt that the HGE set is far more manageable for the analysis of electromagnetic fields propagating in longitudinally non-uniform optical structures. The HGE set has thus been considered as the most convenient basis set for the model developed in this thesis.

Appendix 2.1

Derivation of the Hermite-Gauss Beams

This appendix shows the derivation of the fundamental Gaussian Beam from the paraxial wave equation

$$\partial_x^2 f - 2ip\partial_z f = 0 \qquad [1.7.6]$$

[3], [4]. Assume an ansatz for equation (1.7.6) of the following form:

$$f(x,z) = \sqrt{A(z)}\exp\left(-ip\frac{x^2}{2q(z)}\right) \qquad (A2.1 - 1)$$

where A(z) and q(z) are two unknown functions of z. [This assumption follows from the experimental observation that laser beams show an intensity transverse profile of the type $e^{-2\frac{x^2+y^2}{w^2}}$. In the following derivation, however, the two-dimensional solution in x and z is considered.]

By substitution of (A2.1 -1) into equation (1.7.6) obtain

$$\left[-ip\left(\frac{1}{A}\frac{dA}{dz}+\frac{1}{q}\right)-\left(1-\frac{dq}{dz}\right)\frac{p^2x^2}{q^2}\right]\sqrt{A}\exp\left(-ip\frac{x^2}{2q}\right)=0 \qquad (A2.1 - 2)$$

Equation (A2.1 - 2) must be valid for all x, hence it must be assumed that

$$\frac{dq(z)}{dz}=1 \qquad (A2.1 - 3)$$

and

$$\frac{dA(z)}{dz}=-\frac{A(z)}{q(z)} \qquad (A2.1 - 4)$$

The solutions of (A2.1 - 3) and (A2.1 - 4) are, respectively,

$$q(z) = q_o + z - z_o \qquad (A2.1 - 5)$$

and

$$\frac{A(z)}{A_o}=\frac{q_o}{q(z)} \qquad (A2.1 - 6)$$

Hence the sought beam-like solution takes the form

$$f(x,z) = \sqrt{\frac{A_o q_o}{q(z)}}\exp\left(-ip\frac{x^2}{2q(z)}\right) \qquad (A2.1 - 7)$$

with q(z) specified by (A2.1 - 5). In the general case q(z) is complex, and hence it is convenient to write

$$\frac{1}{q(z)} = \frac{1}{q_1(z)} - i\frac{1}{q_2(z)} \qquad \text{(A2.1 - 8)}$$

where $q_1(z) = R(z)$ is defined as the radius of curvature of the beam and $q_2(z) = \pi w^2(z)$ with w(z) the Gaussian spot size, since, by so doing, the solution resembles the experimental observation mentioned above. A similar derivation applies for the higher order Hermite-Gauss Beams, [3], [4].

References

[1] M. Alonso, E. J. Finn, *Fundamental University Physics, vol. III, Quantum and Statistical Physics*, Addison-Wesley Publishing Company (1968)

[2] D. Marcuse, *Light Transmission Optics*, Van Nostrand Reinhold (1982)

[3] P. W. Milonni, J. H. Eberly, *Lasers*, John Wiley and sons (1988)

[4] A. E. Siegman, *Lasers*, University Science Books (1986)

[5] E. Butkov, *Mathematical Physics*, Addison-Wesley Publishing Company (1973)

[6] G. Fox, T. Li, *Resonant Modes in a Maser Interferometer*, Bell System Techonoly Journal, 40, (1961), pp. 453-458

[7] H. Kogelnik, T. Li, *Laser Beams and Resonators*, IEEE Proceedings, 54, (1966), pp. 1312-1329

[8] A. Yariv, *Optical Electronics*, University Science Books (1971)

[9] P. F. Goldsmith, *Quasi-Optical Techniques*, IEEE Proceedings, 80, (1992), pp. 1729-1747

[10] T. Shen, Z. Sun, W. Dou, *Asymmetrical Gauss-Hermite Beam Mode Analysis of Corrugated Pyramidal Rectangular Horn*, IEE Proceedings of Microwave Antennas Propagation, 143, n. 5 (1996), pp. 385-389

[11] S. Withington, J. A. Murphy, *Analysis of Diagonal Horns through Gaussian-Hermite Modes*, IEEE Transactions on Antennas Propagation, 40, n. 2 (1992), pp. 198-206

[12] D. Marcuse, *The effect of the* ∇n^2 *Term on the Modes of an Optical Square-Law Medium*, IEEE Journal of Quantum Electonics, Correspondence, (1973), pp. 958-960

[13] A. K. Ghatak, M. S. Sodha, *Inhomogeneous optical waveguides*, Plenum Press (1977)

[14] A. Sommerfeld, *Partial differential equations in Physics. Lectures on theoretical Physics - vol VI*, Academic Press (1964)

[15] T. Rozzi, M. Mongiardo, *Open dielectric waveguides*, IEE Electromagnetic wave series (1997)

[16] H. A. Haus, *Waves and Fields in Optoelectronics*, Prentice-Hall (1984)

[17] D. Marcuse, *Loss Analysis of Single-Mode Fiber Splices*, The Bell System Technology Journal, 56, n.5 (1977), pp. 703-718

[18] D. Marcuse, *Gaussian Approximation of the Fundamental Modes of Graded-Index Fibers*, Journal of the Optical Society of America, 68, n. 1 (1978), pp. 103-109

[19] I.A.Erteza, J. W. Goodman, *A Scalar Variational Analysis of Rectangular Dielectric Waveguides using Hermite-Gaussian Modal Approximations*, Journal of Lightwave Technology, 13, n.3, (1995), pp. 493-506

[20] E. K. Sharma, M. P. Singh, A. Sharma, *Variational Analysis of Optical Fibers with Loss or Gain*, Optics Letters, 18, n. 24 (1993), pp. 2096-2098

[21] K. Iga, *Fundamental of Laser Optics*, Plenum Press (1988)

Chapter 3

Field analysis in optical dielectric structures: a brief review

The analysis of electromagnetic field propagation in dielectric structures can be broadly distinguished in two categories: i) for longitudinally uniform structures it can be classified as an eigenvalue problem, and ii) for longitudinally non-uniform structures as an initial value problem.

In the first category of problems, the eigensolutions of the wave equation are determined for a specified dielectric distribution. The dielectric structure is in this case a waveguide; the eigensolutions are the modes (eigenfunctions) supported by the waveguide and the corresponding propagation constants (eigenvalues). The solution, however, is analytic only for some particular dielectric distributions.

For the second category, the wave equation, often in the paraxial approximation, is solved given the initial field distribution and the boundary conditions associated with the problem. Field propagation in longitudinally uniform waveguides is usually solved by expanding the (total) field in terms of the local modes, (local mode expansion, LME) [1]. This method has been extensively used for solving field propagation in metal (closed) guides [2]. In open waveguides the discrete set of bound modes is complemented by a continuum set of radiation modes and hence the mode expansion technique has to be modified in order to take the radiation modes into account [3], [4]. However, to simplify the calculations the radiation mode set is often neglected, [5]. Longitudinally non-uniform devices may be analysed by approximating the structure by a series of uniform waveguide sections of correspondingly different width, and by applying the LME at all junctions between any two such sections [1], [6]. Besides the LME technique alternative solution schemes have been proposed for solving field propagation in longitudinally non-uniform structures, [7], [8].

It is customary to distinguish the various propagation methods in the two broad categories of (a) numerical methods used to directly solve the wave equation, e.g., the BPM [9], [10], the Finite-Difference Method [11], the Finite-Element Method [12],

and (b) semi-analytic methods, e.g., the variational and perturbation methods. In the latter category may also be included a variety of expansion methods [1], [6], [13] in which the wave equation is transformed into a system of coupled differential equations that can be then solved by standard techniques. It is pertinent to indicate that the model developed in this thesis is based on [13], but has been substantially extended to enable a wider range of structures to be analysed.

In the following sections some of the solution schemes found in the literature are reviewed to point out differences and similarities with the model developed in this thesis. Results from some of these methods will be used later for comparison with results obtained from the method developed in this thesis.

1. Longitudinally uniform structures: eigenvalue problem

Consider in the following discussion (longitudinally uniform) waveguides for which the scalar wave equation

$$\left[\partial_x^2 + \partial_z^2 + k_o^2\varepsilon\right]E(x,z) = 0 \qquad [1.4.4]$$

gives a sufficiently accurate description of the modes. In most practical cases only the bound modes are of interest since they can be individually excited. In this case equation (1.4.4) can be solved by separation of variables, hence using the expression $E(x, z) = F(x)e^{-i\beta z}$ to represent an individual bound mode obtain from (1.4.4) the eigenvalue equation

$$\frac{d^2F(x)}{dx^2} + k_o^2\varepsilon(x)F(x) = \beta^2F(x) \qquad (3.1.1)$$

in which both F(x) and β are unknown. Many methods have been proposed to solve the eigenvalue problem (3.1.1). Three of these will be described in the following subsections: the perturbation method, the global expansion method and the variational method. The above mentioned techniques have been extensively used in the literature because they are versatile and hence applicable to a range of problems.

1.1 The perturbation method

The eigenvalue problem related to a particular dielectric distribution may in some cases be considered as a perturbation of a similar eigenvalue problem the solution to which is known. Both the eigenvalues and the eigenfunctions of the new problem are

then considered as perturbations of the known solutions. As an example consider the eigenvalue problem (3.1.1) where $\varepsilon(x) = \bar{\varepsilon}(x) + \delta\varepsilon = (\bar{n}(x) + \delta n)^2$, with $\bar{\varepsilon}(x) = (\bar{n}(x))^2$ the unperturbed dielectric distribution. Referring to Appendix 3.1, where the details of the formalism are presented, the unperturbed equation is

$$\frac{d^2\bar{F}(x)}{dx^2} + k_o^2\bar{\varepsilon}(x)\bar{F}(x) = \bar{\beta}^2\bar{F}(x) \qquad (3.1.2)$$

where $\bar{F}(x)$, $\bar{\beta}$ are the unperturbed eigensolutions, while the perturbation is $W = k_o^2\delta\varepsilon$ (W is formally referred to as the perturbation operator).

An important result that is obtained by applying the perturbation technique is the correction to the eigenvalue

$$\beta_n^{(1)} = W_{nn} \qquad (3.1.3)$$

where the expression $W_{nn} = \int_{-\infty}^{+\infty} f_n^{(o)*}(x)\left[Wf_n^{(o)}(x)\right]dx$ is usually referred to as expectation value for the operator W. In some practical problems equation (3.1.3) is the essential correction to accurately describe the solution of the perturbed problem, i.e., in those cases in which $f_n(x) \approx f_n^{(o)}(x)$, [Appendix 3.1].

For example, in [14], waveguides with parabolic dielectric distribution are considered, and only the correction to the propagation constant due to the TE/TM polarisation term is to be determined, i.e., with reference to equation (3.1.2) and to equation (A1.4- 6) of Appendix 1.4, $W = k_o^2\varepsilon\frac{d}{dx}\left[\frac{1}{\varepsilon(x)}\frac{d}{dx}\right]$. A similar approach is used in [15] for the analysis of truncated parabolic and exponential dielectric distributions.

The perturbation method is also useful to study the characteristics of realistic waveguides to account for deviations from the idealised perfect case because of imperfections in the structure, [1], or to account for gain/loss. For example, in [16] a (constant) small imaginary part is introduced in the refractive index of the waveguide, i.e., $\delta n = -i\frac{g}{2k_o}$, where g is the optical intensity gain in the medium, and the perturbation to the propagation constants of TE and TM modes is calculated. In the

context of active optical devices, in [17] the perturbation is introduced by the injection of carriers into the initially passive waveguide.

1.2 The global expansion method

A variety of methods can be classified under the category of global, or total field expansion methods. They all have in common the fact that the total field in the structure is expanded in terms of a complete set of (orthonormal) functions. The choice of the basis function set is characteristic of the particular method. According to this formalism, the modal solution of equation (3.1.1) is considered as the total field and is expressed as the following series expansion:

$$F(x) = \sum_k a_k f_k(x) \tag{3.1.4}$$

with f_k the orthonormal basis functions and the (constant) expansion coefficients are

$$a_k = \int_I f_k^*(x) F(x) dx \tag{3.1.5}$$

where I is the appropriate range of integration. By substitution of (3.1.4) into (3.1.1) a set of coupled equations is obtained that may be solved by various techniques, such as the Galerkin method, [18], [19], the Ritz-Galerkin method, [20], and the collocation method, [13]. In general the advantage of using (modal) field expansion methods is that all eigensolutions are obtained simultaneously. However, not only all bound modes of the waveguide are determined, but also a discrete approximation of the continuum (radiation) modes is typically obtained, [6], [21].

Examples of function sets used in the literature are i) Laguerre-Gauss functions, [22], [23]; ii) Hermite-Gauss functions, [13], [18], [20], [21], [24]; iii) sine (cosine) functions [6], [19]; iv) waveguide modes (local modes - eigenmode expansion) [1], [4], [5].

The choice of the basis set invariably affects the approach to the solution of equation (3.1.1). For example, in [6], [19] sine functions are used as the basis functions, but since they do not individually satisfy the radiation condition typical of open structures, the dielectric waveguide has to be enclosed in a fictitious metal box in order to conveniently solve the eigenproblem. In [13], [21] instead, the Hermite-Gauss (HG) functions form the basis set. The HG functions individually satisfy the radiation condition, which means that, differently from the expansion methods using

sine functions the radiation condition for the total field is implicitly included in the HG expansion. This aspect is considered to be of fundamental importance in the choice of the HG functions as the basis set for the method presented in this thesis.

The expression for the field expansion (3.1.4) may vary in detail according to the chosen basis set. For example, in the case of the local mode expansion, [1], the expansion (3.1.4) should contain an integral contribution due to the continuum spectrum. However, some authors have considered in (3.1.4), as an approximation, only the contribution of the discrete modes, [5], since it leads to very much reduced complexity and computation.

The method described in [24] to analyse field propagation in inhomogeneous media with gain or loss, presents a further refinement with respect to those mentioned above. In the analytic method described in [24] not only the field, equation (3.1.4), but also the complex dielectric distribution is expressed in terms of the HG set, i.e.,

$$\varepsilon(x) = \sum_{k=0}^{\infty} b_k f_k(x) \qquad (3.1.6)$$

By the use of (3.1.6) in conjunction with (3.1.4) it is possible to derive a quasi-analytic formalism to solve equation (3.1.1). This method appears to have potential advantage, but the suitability of it to the type of devices analysed in this thesis will be considered as future work.

1.3 The variational method

The first step in the variational method is to define a functional which is stationary for the exact solution of the physical problem. The optimal approximate solution is then found by minimising the resulting variational expression which for modal solutions represents the modal propagation constant [refer to the description of the variational method used to approximate the fundamental mode of a three-layer slab waveguide by a Gaussian function, Chapter 2, Section 3]. The propagation characteristics of slab waveguides and optical fibres have been extensively studied with this method, [1], [25].

The variational method has also been used to solve waveguides with complex dielectric distributions. For example, in [22] the complex eigenvalue problem is solved by combining the variational method with the field expansion method [discussed in

this chapter, Section 1.2]. The particular feature of the method described in [22] is that the unknown (complex) modal field profile is expanded in terms of Laguerre-Gauss (LG) functions, as

$$F(x) = \sum_{k=0}^{N} a_k L_k(x) \tag{3.1.7}$$

where $L_k(x)$ are the LG expansion functions, a_k the complex expansion coefficients and N is the number of expansion terms. The advantage of using the field expansion (3.1.7) is in the higher accuracy achieved in the approximation of the exact field profile. Another advantage of expanding the field in terms of LG functions is that the set of equations obtained by applying the variational technique to determine both β and the (complex) expansion coefficients a_k used in (3.1.7) can be solved with the matrix formalism.

2. Longitudinally non-uniform structures: initial value problem

Longitudinally non-uniform structures have an important role in two important fields: i) as passive devices, e.g., mode converters where the adiabatically flared structure allows the mode profile to expand or contract, [1], [26], [27], and ii) as active devices, e.g., flared semiconductor optical amplifiers [28] and semiconductor lasers [29], [30] where the flared cross-section lowers the power density inside the device which reduces the possibility of catastrophic optical damage of the facets, while retaining the capability of yielding high power.

Electromagnetic field propagation in dielectric tapered structures is not readily solvable. However, it is possible with the aid of some approximations to quite accurately describe the field in longitudinally non-uniform devices and hence it is possible not only to analyse such structures, but also to improve on the design and performance of some of these devices. Among the various methods that have been proposed in the literature to solve for field propagation in longitudinally uniform and non-uniform structures, the mode matching method and the BPM will be briefly described in the following sections because they are extensively used in the literature. The third method for analysing field propagation that is considered in this section is the total field expansion method which has already been introduced in the context of longitudinally uniform waveguides [Section 1.3]. In this respect reference [13] will be

reviewed in some detail since, to a large extent, it forms the starting point for a part of the modelling scheme that is developed in this thesis.

Consider a passive longitudinally non-uniform dielectric structure defined by the dielectric distribution $\varepsilon(x, z)$. As outlined in Chapter 1, Section 4, field propagation in such a structure is governed by the scalar wave equation

$$\left[\partial_z^2 + \partial_x^2 + k_o^2\varepsilon(x,z)\right]F(x,z) = 0 \qquad [1.4.4]$$

Equation (1.4.4) may, in some cases, be satisfactorily approximated by assuming paraxial propagation, e.g., in weakly guiding structures. The field is thus written as $F(x, z) = f(x, z)\, e^{-ipz}$, where $f(x, z)$ is the slowly varying field profile, p is a suitably chosen constant [Chapter 1, Section 7], and equation (1.4.4) reduces to the paraxial wave equation

$$\partial_x^2 f - 2ip\partial_z f + \left(k_o^2\varepsilon - p^2\right)f = 0 \qquad [1.7.5]$$

2.1 The mode matching method

The mode matching method has been extensively used in the analysis of metal (closed) waveguides, [2]. However, since dielectric waveguides are categorised as 'open' any field distribution at an arbitrary cross-section of the waveguide should be expressed as the summation of the contributions of both the guided (discrete) and the radiation (continuous) local modes, [1], [31]. In the immediately following discussion, however, it will be assumed that the contribution of the radiation modes may be neglected which allows for an analysis formalism with much reduced complexity, [5].

The longitudinally non-uniform structure is approximated by a series of uniform (slab) waveguide sections of varying width, Fig. 3.2.1. Each section can then be considered as a local waveguide.

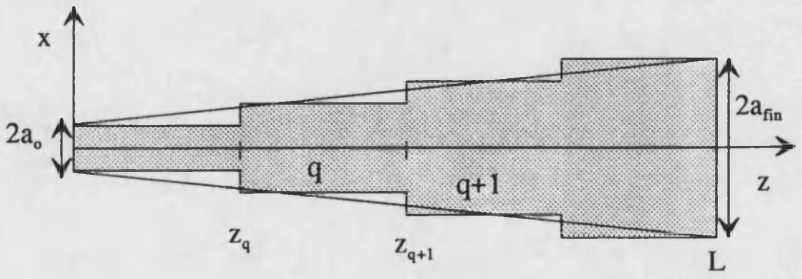


Fig. 3.2.1: Tapered structure approximated by a series of waveguide slabs (e.g., sections q and q+1). Input and output widths: $2a_o$ and $2a_{fin}$, respectively; length of the device: L.

With reference to Fig. 3.2.1, the definition of field matching is reduced to

$$F_{q-1}(x, z = z_q) = F_q(x, z = z_q) \tag{3.2.1}$$

which is a scalar field matching and is valid for weakly guiding structures. The field at the beginning of section q is expressed as

$$F_q(x, z = z_q) = \sum_{k=1}^{N_q} a_{k(q)} f_{k(q)}(x) \tag{3.2.2}$$

where N_q represents the number of bound modes in the uniform section q, $f_{k(q)}$ the bound mode functions of section q, and $a_{k(q)}$ are the (constant) expansion coefficients determined by

$$a_{k(q)} = \int_{-\infty}^{+\infty} f_{k(q)}(x) F_q(x, z = z_q) dx \tag{3.2.3}$$

The (initial) field $F_q(x, z = z_q)$ is then propagated along the uniform section to yield

$$F_q(x, z) = \sum_{k=1}^{N_q} a_{k(p)} f_{k(q)} e^{-i\beta_{k(q)}(z - z_q)} \tag{3.2.4}$$

where $\beta_{k(q)}$ are the propagation constants of the modes in section q. F(x,z) can thus be obtained at $z = z_{q+1}$ and the same procedure is repeated at each junction.

The advantage of this method is that it has a simple and intuitive interpretation in terms of the local modes. However, in general it is not possible to limit the expansion to only the guided modes. The general field must be represented by the complete set comprising of the bound and the radiation modes. Hence equation (3.2.1) must include the continuum spectrum but this adds very significantly to the complexity of the computation, not least because of the numerical integration.

2.2 The beam propagation method

The numerical BPM was introduced in the early 1970s to solve underwater acoustic problems, [9]. It was later applied to analyse fibre optics structures since it was recognised that the same approximations used for underwater acoustic analysis were also valid for the analysis of (weak) dielectric structures, [32]. Various versions of the Beam Propagation Method (BPM) have thus been proposed to solve for field propagation problems ranging from the scalar paraxial to the full vectorial analysis. To simplify the approach to this method the simple version of the BPM based on the

scalar paraxial wave equation, [9], is discussed below. The version of the BPM described here is a combination of the eigenfunction expansion method and the perturbation method.

The BPM assumes that the longitudinal non-uniformity of the structure is small so that the refractive index distribution may be written as

$$n(x,z) = \bar{n}(x) + \Delta n(x,z) \tag{3.2.5}$$

with Δn small. The wave equation can thus be considered in the scalar form, (1.4.4). It is also assumed that the solutions to the scalar wave equation with $n(x) = \bar{n}(x)$ are known as $f_k(x)$ (eigenfunctions) and β_k (eigenvalues). The choice of $\bar{n}(x)$ has to satisfy two criteria: i) that Δn is small, and ii) that the resulting numerical procedure is fast and stable, [9]. It has been found that it is convenient to choose $\bar{n}(x)$ to be constant which means that the unperturbed problem is that of free-space propagation - the eigenfunctions $f_k(x)$ are thus plane waves.

The details of the formalism have been outlined in Appendix 3.2. However, the most relevant result is that, following some assumptions, the first order approximation to the field due to the non-uniformity $\Delta n(x, z)$ is in the phase, $e^{-ik_o \Delta n \Delta z}$. It is worth noting that this correction typically describes the focusing effect of thin lenses. Therefore, in the first approximation the field propagating in the longitudinally non-uniform medium is given by

$$F(x,z) = E(x,z)e^{-ik_o \Delta n \Delta z} \tag{3.2.6}$$

where E(x,z) is the field propagating in the unperturbed medium ($\bar{n}$) and Δz is the propagation step in the longitudinal direction [Appendix 3.2]. The simplified version of the BPM algorithm [9] reviewed here is subject to a few conditions that limit the range of Δn and ∇n that can be considered; however, improved versions of the BPM are available in the literature so that the number of such constraints is reduced, [33].

2.3 The total field propagation scheme

This section presents the global expansion scheme with the particular use of the Hermite-Gauss (HG) functions, as was suggested in [13], as being suitable for the analysis of longitudinally (weakly) non-uniform dielectric structures. The applicability

of the HG expansion scheme also to the analysis of periodic structures and nonlinear pulse propagation in optical fibres is presented in [34].

The advantages of the total field expansion method based on the HG functions are that

i) it is applicable to arbitrary dielectric profiles and arbitrary waveguide shapes;

ii) it involves no inherent approximations;

iii) it solves for the total field;

iv) it naturally leads to the use of the (numerical) collocation method by means of which the scalar partial differential equation can be easily reduced to a set of coupled ordinary differential equations.

The solution of equation (1.4.4), F(x, z), is sought as a linear combination of HG functions

$$F(x,z) = \sum_k A_k(z) f_k(x) \tag{3.2.7}$$

where $A_k(z)$ are the z-dependent expansion coefficients, and $f_k(x)$ are the HG functions. Substitution of (3.2.7) into (1.4.4) or (1.7.5) yields the set of coupled equations for the field expansion coefficients.

In [13] the HG expansion scheme has been compared with the Galerkin method for the eigenvalue problem related to various dielectric distributions, obtaining excellent agreement. Further, it has also been compared with the BPM for the solution of field propagation problems, finding the HG expansion scheme to be more convenient for both accuracy and computational efforts. However, the range of cases tested in [13] have been limited and that shortcoming has also been addressed in this thesis.

Summary

The two categories of electromagnetic field analysis problems, the eigenvalue and the initial value (propagation) problems, have been discussed reviewing some of the most commonly used methods of solution.

Attention has been drawn in this chapter to semi-analytic methods (the perturbation method, the global expansion method and the variational method) because the intention has been to emphasise possible connections between these methods and the solution scheme developed in the remainder of this thesis.

For the category of field propagation (initial value) problems in longitudinally non-uniform structures three methods have been reviewed. The first method is the quasi-analytic method of mode matching. The reviewed beam propagation scheme is a simplified version of the Beam Propagation Method (BPM). This version of the BPM solves for field propagation in (weakly) non-uniform structures using a perturbation technique where the perturbed solution is expanded in terms of plane waves. The third method is a more general scheme typically referred to as global expansion method. In this method the total field propagating in the structure is expanded in terms of a suitably chosen basis set. In fact, the field analysis developed in this thesis is of this third type and uses the Hermite-Gauss functions as the basis set.

Appendix 3.1

First order perturbation technique

For the following derivation reference is made to [35]. Consider an eigenvalue problem of the type

$$(H+W)f_n(x)=\beta_n f_n(x) \qquad \text{(A3.1 - 1)}$$

where H and W are two linear, hermitian operators, $f_n(x)$ and β_n the eigenfunction and corresponding eigenvalue. The operator W is considered as a perturbation of the operator H. The eigenvalue problem associated with H is

$$Hf_n^{(o)}(x)=\beta_n^{(o)}f_n^{(o)}(x) \qquad \text{(A3.1 - 2)}$$

$f_n^{(o)}$ and $\beta_n^{(o)}$ are the eigenfunctions and eigenvalues of H satisfying the condition $\beta_m^{(o)} \neq \beta_k^{(o)}$ for $m \neq k$ (a modified derivation is necessary if the operator H is degenerate). It is convenient to express the perturbation W in terms of a (small) coefficient α as $W = \alpha W^{(1)}$ because it is expected that when $\alpha \to 0$ (the perturbation is zero) the eigenfunctions and eigenvalues of (A3.1 - 1) tend to those of the unperturbed problem (A3.1 - 2). Hence, the perturbed eigenfunctions can be expanded in terms of the powers of the parameter α, and similarly for the eigenvalues:

$$f_n(x)=\sum_k \alpha^k \tilde{f}_n^{(k)}(x) \qquad \text{(A3.1 - 3)}$$

and

$$\beta_n=\sum_k \alpha^k \tilde{\beta}_n^{(k)} \qquad \text{(A3.1 - 4)}$$

Hence in the limit $\alpha \to 0$ the perturbed solution tends to the unperturbed one, i.e., $\forall n \;\; f_n(x) \xrightarrow[\alpha\to 0]{} f_n^{(o)}(x)$ and $\beta_n \xrightarrow[\alpha\to 0]{} \beta_n^{(o)}$. For simplicity put $f_n^{(i)}(x)=\alpha^i \tilde{f}_n^{(i)}(x)$ and $\beta_n^{(i)}=\alpha^i \tilde{\beta}_n^{(i)}$ (i = 1, 2, ...) and consider only the first order corrections to $f_n^{(o)}$ and $\beta_n^{(o)}$ so that equations (A3.1 - 3) - (A3.1 - 4) become

$$f_n(x)=f_n^{(o)}(x)+f_n^{(1)}(x) \qquad \text{(A3.1 - 5)}$$

and

$$\beta_n=\beta_n^{(o)}+\beta_n^{(1)} \qquad \text{(A3.1 - 6)}$$

Substituting equations (A3.1 - 5) - (A3.1 - 6) into (A3.1 - 1) obtain

$$Hf_n^{(1)}(x)+Wf_n^{(o)}(x)=\beta_n^{(o)}f_n^{(1)}(x)+\beta_n^{(1)}f_n^{(o)}(x) \qquad \text{(A3.1 - 7)}$$

Multiplying by $f_n^{(o)}(x)$ and integrating over $(-\infty,+\infty)$ equation (A3.1 - 7) becomes

$$\int f_n^{(o)*}\left(Hf_n^{(1)}\right)dx + \int f_n^{(o)*}\left(Wf_n^{(o)}\right)dx = \beta_n^{(o)}\int f_n^{(o)*}f_n^{(1)}dx + \beta_n^{(1)}\int f_n^{(o)*}f_n^{(o)}dx \quad \text{(A3.1 - 8)}$$

Since the operators H and W are linear and hermitian (A3.1 - 8) can be reduced to

$$\beta_n^{(1)} = W_{nn} \quad \text{(A3.1 - 9)}$$

where $W_{nn} = \int_{-\infty}^{+\infty} f_n^{(o)}(x)\left(Wf_n^{(o)}(x)\right)dx$ is referred to as expectation value for the operator W. Equation (A3.1 - 9) is a very important result since it determines the first order correction to the unperturbed eigenvalue due to the perturbation W on H.

The derivation that follows is necessary to determine the (first order) correction to the eigenfunctions. For this purpose it is essential that the unperturbed eigenfunctions $f_n^{(o)}(x)$ form a complete set of functions. In this case, then, $f_n^{(1)}(x)$ can be expanded in terms of such a basis set, obtaining

$$f_n^{(1)}(x) = \sum_q c_{nq} f_q^{(o)}(x) \quad \text{(A3.1 - 10)}$$

Using equation (A3.1 - 10) in (A3.1 - 7), multiplying by $f_p^{(o)}(x)$ $(p \neq n)$ and integrating over $(-\infty,+\infty)$ gives

$$\sum_q c_{nq}\int f_p^{(o)*}\left(Hf_q^{(o)}(x)\right)dx + W_{pn} = \beta_n^{(o)}c_{np} \quad \text{(A3.1 - 11)}$$

where W_{pn} depicts the matrix element of the operator W between vectors labelled p and n, i.e., $W_{pn} = \int_{-\infty}^{+\infty} f_p^{(o)*}(x)\left(Wf_n^{(o)}(x)\right)dx$. Thus from equation (A3.1 - 11) obtain the unknown coefficients c_{np} of expansion (A3.1 - 10) making use of the linearity property of H:

$$c_{np} = \frac{W_{pn}}{\beta_n^{(o)} - \beta_p^{(o)}} \quad \text{(A3.1 - 12)}$$

However, equation (A3.1 - 12) gives all the coefficients c_{np} with $p \neq n$. The last coefficient c_{nn} is obtained by imposing the normalisation condition on the perturbed eigenfunctions. This condition leads to

$$c_{nn} + c_{nn}^* = 0 \quad \text{(A3.1 - 13)}$$

which implies that if the functional space is real, $c_{nn} = 0$; if it is complex the expansion coefficients are purely imaginary since (A3.1 - 13) enforces $\mathrm{Re}(c_{nn}) = 0$, but arbitrary since $\mathrm{Im}(c_{nn})$ remains undetermined.

In cases in which the first order corrections are not sufficiently accurate, it is necessary to consider higher order corrections for the solution of the perturbed problem, [35].

Appendix 3.2

BPM: derivation of the correction term e^{Γ}

In the following appendix, a detailed discussion of the BPM scheme, [9], is presented. Field propagation in a weakly non-uniform medium is described by the scalar wave equation

$$\left[\partial_x^2 + \partial_z^2 + k_o^2 n^2(x,z)\right]F(x,z) = 0 \qquad [1.4.4]$$

The non-uniform medium is specified by the following refractive index distribution

$$n(x,z) = \bar{n}(x) + \Delta n(x,z) \qquad (A3.2 - 1)$$

with $\bar{n}$ constant and $\Delta n(x,z)$ a small perturbation of $\bar{n}$. The unperturbed differential equation related to $\bar{n}$ is

$$\frac{d^2 f}{dx^2} + \left(k_o^2 \bar{n}^2 - \beta^2\right) f(x) = 0 \qquad (A3.2 - 2)$$

with f(x) the unperturbed eigenfunctions (plane waves), and β the unperturbed and eigenvalues. The field at a specified longitudinal position, z, in the non-uniform medium is assumed to be accurately described by

$$F(x,z) = E(x,z)\, e^{\Gamma(x,z)} \qquad (A3.2 - 3)$$

where E(x,z) is the field propagating in the uniform medium ($\bar{n}$), which can be expressed in terms of the unperturbed eigenfunctions as a Fourier Transform; and $e^{\Gamma(x,z)}$ is the (multiplicative) correction term due to the non-uniformity $\Delta n(x,z)$.

In the present formulation only the forward travelling components are considered since it is supposed that the contribution of the reverse travelling components to the forward travelling beam is negligible. The individual components of E(x,z) in (A3.2-3) are thus propagated in the homogeneous medium and hence the field can be recomposed at any longitudinal position.

The multiplicative correction $e^{\Gamma(x,z)}$ is determined using the perturbation technique. Thus, substituting equation (A3.2 - 3) into the scalar wave equation, using (A3.2 - 2) and retaining only the first order correction terms, obtain

$$E\left(\partial_x^2\Gamma + \partial_z^2\Gamma\right) + 2\left(\partial_x\Gamma\partial_x E + \partial_z\Gamma\partial_z E\right) + k_o^2\left(n^2(x,z) - \bar{n}^2\right)E = 0 \qquad (A3.2 - 4)$$

Since the propagation step Δz can be made arbitrarily small, the exponent $\Gamma(x,z)$ of the correction term can be expanded in a Taylor series as

$$\Gamma(x,z) = \sum_{s=1}^{\infty} A_s(x) z^s \qquad \text{(A3.2 - 5)}$$

Substituting (A3.2 - 5) into (A3.2 - 4) and making use of equation (A3.2 - 2) obtain the set of coupled equations for the expansion coefficients $A_s(x)$ that define $\Gamma(x,z)$ in (A3.2 - 5). The first coefficient (first order correction) is given by

$$A_1^2 + \left(\frac{2}{E}\partial_z E\right)A_1 + k_o^2\left(n^2(x,z) - \bar{n}^2\right) + 2A_2 = 0 \qquad \text{(A3.2 - 6)}$$

It is possible to simplify equation (A3.2 - 6) assuming that

$$A_2 << \frac{k^2\left(n^2(x,z) - \bar{n}^2\right)}{2} \qquad \text{(A3.2 - 7)}$$

Using thus the approximation (A3.2 - 7) the first coefficient becomes

$$A_1 = \frac{1}{E}\partial_z E\left(-1 + \sqrt{1 - \frac{k^2\left(n^2(x,z) - \bar{n}^2\right)}{\left(\frac{1}{E}\partial_z E\right)^2}}\right) \qquad \text{(A3.2 - 8)}$$

To further simplify the expression for A_1 it is convenient to assume that

$$\frac{k_o^2\left(n^2(x,z) - \bar{n}^2\right)}{\left(\frac{1}{E}\partial_z E\right)^2} << 1 \qquad \text{(A3.2 - 9)}$$

which is equivalent to assuming paraxial propagation. Finally, the expression for A_1 is thus

$$A_1 = \frac{k_o^2\left(n^2(x,z) - \bar{n}^2\right)}{2}\frac{E}{\partial_z E} \qquad \text{(A3.2 - 10)}$$

For each individual eigenfunction (plane wave) of the unperturbed problem (A3.2 -2), the first order correction term is in the phase. The first coefficient for $\Gamma(x,z)$ is

$$A_1 = -ik_o\Delta n \qquad \text{(A3.2 - 11)}$$

and hence the correction term defined in (A3.2 - 3) is

$$e^{\Gamma} = e^{-ik_o\Delta n\Delta z} \qquad \text{(A3.2 - 12)}$$

which describes the focusing effect of the (Δz) layer of non-uniform (Δn) medium on the unperturbed component of the field.

References

[1] D. Marcuse, *Light transmission optics*, Van Nostrand Reinhold (1982)

[2] F. Sporelder, H. G. Unger, *Waveguide tapers transitions and couplers*, Peter Peregrinus Ltd. (1979)

[3] V. V. Shevchenko, *Continuous transitions in open waveguides*, The Golem Press (1971)

[4] T. Rozzi, M. Mongiardo, *Open electromagnetic waveguides*, IEE Electromagnetic Waves Series (1997)

[5] I. Middlemast, J. Sarma, P. S. Spencer, *Characteristics of tapered rib-waveguides for high-power semiconductor optical sources*, IEE Proceedings-Optoelectronics, 144 (1997), pp. 8-13

[6] C. H. Henry, Y. Shami, *Analysis of mode propagation in optical waveguide devices by Fourier expansion*, IEEE J. Q. E. 27 (1991), pp. 523-530

[7] S. M. Saad, *Review of numerical methods for the analysis of arbitrarily-shaped microwave and optical dielectric waveguides*, IEEE Trans. on Microwave Theory and Technology, MTT-33 (1985), p. 894-899

[8] K. S. Chiang, *Review of numerical and approximate methods for the modal analysis of general dielectric waveguides*, Optical and Quantum Electronics 26 (1994), pp. S113-S134

[9] P. E. Lagasse, R. Baets, *The Beam Propagation Method in integrated optics*, in Hybrid Formulation of Wave Propagation and Scattering, ed. L. B. Felsen, NATO ASI Series, Martinus Nijhoff Publisher (1984)

[10] J. Van Roey, J. van der Donk, P. E. Lagasse, *Beam propagation method: analysis and assessment*, Journal of the Optical Society of America, 71 (1981), pp. 803-810

[11] R. Accornero, M. Artiglia, G. Coppa, G. L. P. Di Vita, M. Potenza, P. Ravetto, *Finite difference methods for the analysis of integrated optical waveguides*, Electronics Letters, 16 (1990), pp. 1959-1960

[12] J. Brian Davies, *Finite element analysis of waveguides and cavities - a review*, IEEE Transactions on Magnetics, 29 (1993), pp.1578-1583

[13] A. Sharma, S. Banerjee, *Method for propagation of total fields or beams through optical waveguides*, Optics Letters, 14 (1989), pp. 96-98

[14] D. Marcuse, *The effect of the ∇n^2 term on the modes of an optical square-law medium*, IEEE Journal of Quantum Electronics, correspondence QE-9 (1973), pp. 958-960

[15] A. Kumar, K. Thyagarajan, A. K. Ghatak, *Modes in inhomogeneous slab waveguides*, IEEE Journal of Quantum Electronics, QE-10 (1974), pp 902-904

[16] W. H. Weber, S. L. McCarthy, G. W. Ford, *Perturbation theory applied to gain or loss in optical waveguides*, Applied Optics, 13 (1974), pp. 715-716

[17] N. S. Brooks, J. Sarma, I. Middlemast, *A compact model of tapered geometry semiconductor optical devices*, Semiconductor and Integrated Optoelectronic Conference (SIOE '96), paper 45, abstract, Technical Digest, Cardiff, April 1996

[18] A. Weisshaar, J. Li, R. Gallawa, I. C. Goyal, *Vector and quasi vector solutions for optical waveguide modes using efficient Galerkin' s method with Hermite-Gauss basis functions*, Journal of Lightwave Technology, 13 (1995), pp. 1795-1800

[19] D. Marcuse, *Solution of the vector wave equation for general dielectric waveguides by Galerkin method*, IEEE Journal of Quantum Electronic, 28 (1992), pp. 459-465

[20] J. P. Meunier, J. Pigeon, J. N. Massot, *A numerical technique for the determination of propagation characteristics of inhomogeneous planar optical waveguides*, Optical and Quantum Electronics, 15 (1983), pp. 77-85

[21] R. L. Gallawa, I. C. Goyal, Y. Tu, A. K. Ghatak, *Optical waveguide modes: an approximate solution using Galerkin' s method with Hermite-Gauss basis functions*, IEEE J. Q. E. 27 (1991), pp. 518-522

[22] E. K. Sharma, M. P. Singh, A. Sharma, *Variational analysis of optical fibers with loss or gain*, Optics Letters, 18 (1993), pp. 2096-2098

[23] J. P. Meunier, J. Pigeon, J. N. Massot, *A general approach to the numerical determination of modal propagation constants and field distributions of optical fibers*, Optical and Quantum Electronics, 13 (1981) pp. 71-83

[24] R. McDuff, *Matrix method for beam propagation using Hermite polynomials*, Applied Optics, 29 (1990), pp. 802-808

[25] D. Marcuse, *Gaussian approximation of the fundamental modes of graded-index fibers*, Journal of the Optical Society of America, 68 (1978), pp.103-109

[26] A. R. Nelson, *Coupling optical waveguides by tapers*, Applied Optics, 14 (1975), pp. 3012-3015

[27] A. F. Milton, W. K. Burns, *Mode coupling in optical waveguide horns*, IEEE Journal of Quantum Electronics, QE-13 (1977), pp. 828-835

[28] S. Obrien, D. F. Welch, R. A. Parke, D. Mehuys, K. Dzurko, R. J. Lang, R. Waarts, D. Scifres, *Operating characteristics of a high-power monolithically integrated flared amplifier master oscillator power-amplifier*, IEEE Journal of Quantum Electronics, 29 (1993), pp. 2052-2057

[29] K. A. Williams, J. Sarma, I. H. White, R. V. Penty, I. Middlemast, T. Ryan, F. R. Laughton, J. S. roberts, *Q-switched bow-tie lasers for high-energy picosecond pulse generation*, Electronics Letters, 30 (1994), pp. 320-321

[30] D. Mehuys, S. Obrien, R. J. Lang, A. Hardy, D. F. Welch, *5W, diffraction limited, tapered-stripe unstable resonator semiconductor laser*, Electonics Letters, 30 (1994), pp. 1855-1856

[31] W. C. Chew, *Waves and fields in inhomogeneous media*, IEEE Press (series on electromagnetic waves), (1995)

[32] M. D. Feit, J. A. Fleck, *Ligth propagation in graded-index optical fibers*, Journal of Applied Optics, 17 (1978), pp. 3390-3398

[33] C. Vassallo, F. Collino, *Highly efficient absorbing boundary conditions for the beam propagation method*, Journal of Lightwave Technology, 14 (1996), pp. 1570-1577

[34] A. Sharma, *Collocation method for wave propagation through optical waveguiding structures*, in Progress in Electromagnetics Research, Electromagnetic Waves PIER 11, J. A. Kong ed., EMW Publishing (1995)

[35] E. Butkov, *Mathematical physics*, Addison-Wesley Publishing Company (1973)

Chapter 4

Total field analysis using the Hermite-Gauss function expansion method

The motivation for the work in this thesis has been to develop a convenient and efficient scheme for modelling a range of optical devices of current interest that may be classified as dielectric structures that propagate weakly diffracting signals. The essential field problem is to solve for weakly diffracting fields in inhomogeneous media. The Local Mode Expansion (LME) is possible in principle, but the difficulties of including and computing the radiation modes are well known. It is proposed and demonstrated in this chapter that the Hermite-Gauss (HG) function expansion provides for a convenient alternative method of analysis that reduces the difficulties associated with the LME. The same method will be used in the following chapters to solve also the carrier diffusion equation. This will then lead to the development of a compact and self-consistent formulation for the solution of both the field and the carrier profile in active optical devices by the same method.

The main concern in this chapter is the solution of the wave equation in passive open structures using a total field expansion technique of the type described in brief in Chapter 3, Sections 1.2 and 2.3, based on the HG function set. As outlined in Chapter 3, total field expansion methods are versatile and readily applicable to a variety of problems. Here the HG expansion method has been used for the solution of field problems not only in (longitudinally uniform) waveguides but also in longitudinally non-uniform structures of arbitrary shape and dielectric distribution.

This chapter is arranged as follows. The choice of the HG set as the basis on which to expand the optical field is justified in Section 1. The analytic formalism for the HG expansion method is described in Section 2, while the discretisation (analytic) procedure to numerically solve the resulting set of equations is presented in Section 3. In Section 4 the numerical collocation method is associated with the HG expansion method to form the Hermite-Gauss Collocation Method (HGCM) that is used to obtain the numerical solution of field propagation problems. The correspondence of the HGCM formalism with the analytic method as discussed in Section 3 is then

established. Finally, the choice of the (arbitrary) width parameter for the HG functions is discussed in Section 5.

1. Basis function set

The choice of the most suitable basis functions set is dictated by the problem to be solved. Each physical problem is defined by the boundary conditions, and hence the solution must satisfy not only the differential equation associated with the problem, but also the specified boundary condition. For (weakly guiding) open structures the boundary condition associated with the electromagnetic (scalar) field is the radiation condition which imposes that the field be continuous and vanish at infinity, such that the power carried by the electromagnetic field is finite. Electromagnetic (optical) fields are represented by 'well-behaved' functions in the (vector) space of square integrable functions defined on the infinite domain, $L^2((-\infty,+\infty))$. The solutions of interest F(x,z) are thus characterised by the following property

$$\int_{-\infty}^{+\infty}|F(x)|dx < \infty \qquad \forall z \qquad (4.1.1)$$

In the functional space of interest, $L^2((-\infty,+\infty))$, it is possible to specify various sets of functions which are characterised by the same property of completeness. Each such function set is defined as a basis set and it can be demonstrated that any element of $L^2((-\infty,+\infty))$ can be represented by a linear combination of basis functions, [1], [2]. The Fourier transform is one well-known example of function expansion on a basis set of $L^2((-\infty,+\infty))$, but other examples exist such as the Hermite-Gauss or Laguerre-Gauss function expansion, which have been mentioned in Chapter 3, Section 1.2.

For the analysis of electromagnetic field propagation in passive weakly guiding dielectric structures, a convenient and suitable basis set has been found in the Hermite-Gauss functions. The following discussion on the most relevant properties of the HG set provides a justification for this choice:

1) the HG functions constitute a complete set of orthogonal functions in the space of square integrable functions on the interval $(-\infty,+\infty)$, which covers the solutions of the field problems of interest. The completeness property follows from the fact that the HG functions are the solutions of a Sturm-Liouville differential equation [Chapter 2, Section 2], [1], [2];

2) the HG functions individually satisfy the radiation condition for the total field in open dielectric structures;

3) the HG functions form a discrete set. This property is very important in the simplification of the mathematical formalism associated with the field analysis problems. Since the HG functions form a complete set the total optical field is accurately represented by the expansion in terms of HG functions, i.e., the contributions of both the bound and the radiation (local) modes of the open structures are taken into account at the same time under the series summation. An individual radiation mode, which is a practically unrealisable solution, cannot be described by a summation of HG functions. This limitation is however only apparent since the intention is to solve for the total field;

4) the HG functions represent a good approximation to the bound modes of three-layer symmetric slab dielectric waveguides [refer to Section 3, Chapter 2]. However note that for a rigorous description of each individual waveguide mode the totality of HG functions should be considered.

Because of the properties listed above the complete discrete set of HG functions constitutes a convenient basis set for field analysis in open structures. In particular the fact that they satisfy the radiation boundary condition associated with open structures makes it possible to incorporate the radiation boundary conditions for electromagnetic field problems (in open structures) implicitly in the formalism. Finally note that since the HG functions are the eigenmodes of the (infinitely extended) parabolic dielectric distribution, the HG field expansion may be interpreted also in the framework of perturbation methods [refer to the following Section 2.1].

2. Formalism for the HG expansion method

In the following two subsections the eigenvalue problem for (longitudinally uniform) waveguides and the field propagation (initial value) problem in passive dielectric structures are analysed by using the total field expansion method based on the Hermite-Gauss functions.

2.1 The eigenvalue problem

Longitudinally uniform waveguides are analysed in this section so that the dielectric distribution $\varepsilon(x,z) = \varepsilon(x)$ is considered. Assuming mode solutions of the type (1.6.1), $E_m(x, z) = F_m(x)\exp(-i\beta_m z)$, the wave equation takes the form of an eigenvalue equation

$$\frac{d^2F_m(x)}{dx^2} + k_{x,m}^2(x)F_m(x) = 0 \qquad [1.10.3]$$

where $F_m(x)$ is the mode field profile (eigenfunction) and $k_{x,m}^2(x) = k_o^2\varepsilon(x) - \beta_m^2$, with β_m the corresponding propagation constant (eigenvalue).

The field expansion method of analysis may applied here by expanding a mode function in terms of the complete set of HG basis functions so that for the m-th mode

$$F_m(x) = \sum_{k=0}^{\infty} a_k^{(m)} f_k(x) \qquad (4.2.1)$$

where $a_k^{(m)}$ are the (constant) expansion coefficients. By substituting equation (4.2.1) into the eigenvalue equation (1.10.3) making use of the properties of the basis functions, equation (2.2.6) and (2.2.7), obtain i) the equation for evaluating β_m, as

$$\beta_m^2 = \int_{-\infty}^{+\infty} f_m^*(x)\left[k_o^2\varepsilon(x) + x^2 - (2m+1)\right]f_m(x)dx \qquad (4.2.2)$$

where m is the mode number;

and ii) the set of coupled equations for evaluating the expansion coefficients, $a_k^{(m)}$, which are needed to reconstruct the modal field

$$\sum_k a_k^{(m)} \int_{-\infty}^{+\infty} f_m^*(x)\left[k_o^2\varepsilon(x) + x^2\right]f_k(x)dx = 0 \qquad (4.2.3)$$

where $k \neq m$, [Appendix 4.1.]

Two aspects of the above results may be highlighted. The first is that by writing equation (4.2.2) in the form

$$\beta_m^2 - \tilde{\beta}_m^2 = k_o^2 \int_{-\infty}^{+\infty} f_m^*(x)\left[\varepsilon(x) - \tilde{\varepsilon}(x)\right]f_m(x)dx \qquad (4.2.4)$$

where $\tilde{\varepsilon}(x) = -\frac{x^2}{k_o^2} + b$ is the dielectric distribution of a quadratic medium, and

$\tilde{\beta}_m = \sqrt{k_o^2 b - (2m+1)}$ the propagation constant for the corresponding modes, the

solution β_m corresponding to $\varepsilon(x)$ may be recognised as a perturbation solution, [refer to Section 1.1, Chapter 3].

The second point is that equation (4.2.3) gives the expansion coefficients a_k only for $k \neq m$. The coefficient, $a_m^{(m)}$, is determined by normalising each eigenfunction, i.e., by enforcing that

$$a_m^{(m)} = \sqrt{1 - \sum_{k \neq m} \left(a_k^{(m)}\right)^2} \tag{4.2.5}$$

With the last coefficient specified by (4.2.5) the waveguide mode of order m can be recomposed as the HG series (4.2.1), with the propagation constant given by (4.2.4).

2.1.1 Complex media

The HG expansion method has been applied also to waveguides with gain or loss, i.e., with complex dielectric distribution. The (complex) m-th mode function may be written as

$$F_m(x) = F_{m,R}(x) + iF_{m,I}(x) \tag{4.2.6}$$

Expanding $F_m(x)$ in terms of HG functions implies that both the real, $F_{m,R}(x)$, and the imaginary part, $F_{m,I}(x)$, of the mode function can be expressed in terms of the HG set. This leads to the expression

$$F_m(x) = \sum_k a_k' f_k(x) + i\sum_k a_k'' f_k(x) = \sum_k (a_k' + ia_k'') f_k(x) = \sum_k a_k f_k(x) \tag{4.2.7}$$

where now $a_k = a_k' + ia_k''$ are the complex expansion coefficients.

Waveguides with loss or gain are described by non-self-adjoint operators. Although for waveguides with complex dielectric distribution the characteristic equation is formally identical to equation (1.10.3), the operator $L = \frac{d^2}{dx^2} + k_o^2\varepsilon(x)$ is now complex. This implies that the eigenmodes, $u_n(x)$, of the non-Hermitian operator L which describes the complex system do not form a complete set of orthogonal functions. In fact, the eigenmodes of complex systems are biorthogonal, i.e., they are orthogonal to a set of functions which are the eigensolutions of the adjoint operator, $L^\dagger$, of L.

The eigenmodes of a complex parabolic dielectric distribution are still HG functions, but with complex argument, i.e., they are defined by a complex width

parameter w_o, [3]. An electromagnetic field in a complex system should be expanded in terms of the Complex Hermite-Gauss (CHG) functions. It is recognised that equation (4.2.7) is thus an approximated expression for the field, but it is argued that it is sufficiently accurate for media with small imaginary part in the complex refractive index, as also found in [4].

A field expansion scheme that uses the CHG functions as the basis set is presented in [5]. However it is found that the formalism is excessively complicated because of the need to operate with the biorthogonal CHG set.

2.2 The initial value problem

Consider now the problem of field propagation in longitudinally non-uniform, weakly guiding structures described by the dielectric distribution $\varepsilon(x, z)$. The equation to be solved is the scalar wave equation

$$\left[\partial_x^2 + \partial_z^2 + k_o^2\varepsilon(x,z)\right]E(x,z) = 0 \qquad [1.4.4]$$

The method of solution used here is similar to the one described in the previous section, except that now the HG expansion of the total field, E(x, z), involves complex and z-dependent coefficients to account for propagation. The expression (4.2.1) is thus replaced by

$$E(x,z) = \sum_{k=0}^{\infty} B_k(z) f_k(x) \qquad (4.2.8)$$

with $\{f_k(x)\}$ the HG basis functions, (2.2.4), and $B_m(z)$ the expansion coefficients.

Equation (4.2.8) is substituted in the scalar wave equation (1.4.4). Then, as presented in Appendix 4.2, using the properties of the (HG) basis functions, equation (1.4.4) is reduced to a set of coupled ordinary differential equations for the expansion coefficients $B_k(z)$,

$$\frac{d^2B_k(z)}{dz^2} - \frac{2k+1}{w_o^2}B_k(z) + \sum_j B_j(z)Q_{kj}(z) = 0 \qquad (4.2.9)$$

where k = 1, 2, ... and

$$Q_{kj}(z) = \int_{-\infty}^{+\infty} f_k^*(\xi)\left[\left(\frac{\xi}{w_o}\right)^2 + k_o^2\varepsilon(\xi,z)\right]f_j(\xi)d\xi \qquad (4.2.10)$$

is the matrix element, [Appendix 4.2].

2.2.1 Paraxial wave equation

The set of second order coupled differential equations (4.2.9) may be further simplified by making appropriate approximations. For example for field propagation in weakly guiding structures the paraxial approximation is typically used to reduce the complexity of having to solve coupled second order differential equations. In this case, then, the field is written in the form $E(x, z) = F(x, z)e^{-ipz}$, equation (1.6.3). Equation (1.4.4) is reduced to the paraxial wave equation (1.7.5)

$$\partial_x^2 F - 2ip\partial_z F + \left(k_o^2 \varepsilon(x,z) - p^2\right)F = 0 \qquad [1.7.5]$$

which can be solved as an initial value problem. The slowly varying field profile F(x,z) is now expanded in terms of the HG basis functions, (2.2.4), as

$$F(x,z) = \sum_{k=0}^{\infty} b_k(z) f_k(x) \qquad (4.2.11)$$

with $b_k(z)$ the complex z-dependent expansion coefficients. Following the same method used to derive (4.2.9) obtain the set of first order coupled differential equations for the expansion coefficients $b_k(z)$:

$$-2ip\frac{db_k(z)}{dz} - \frac{2k+1}{w_o^2} b_k(z) - \frac{p^2}{w_o^2} b_k(z) + \sum_j b_j(z) Q_{kj}(z) = 0 \qquad (4.2.12)$$

with k = 1, 2, ... and the matrix element Q_{kj} is defined as in (4.2.10).

3. Discretisation

The expansion coefficients used in the total field expansions (4.2.1), (4.2.8) and (4.2.11), derive from the solution of the sets of coupled equations, (4.2.3) - (4.2.9) - (4.2.12), respectively. In this section the analytic orthogonalisation process to solve the above mentioned sets is presented. In the formulation of the propagation scheme presented in this thesis, the analytic method has been replaced by the faster and more convenient collocation numerical method. However, it is useful to point out that the two procedures are inter-connected by the Gaussian Quadrature Formula, [6], as shown below.

Consider in the following discussion the field expansion (4.2.8). For a practical solution only a finite number, M, of expansion terms is used, so that (4.2.8) becomes

$$E_{(M)}(x,z) = \sum_{k=1}^{M} B_k(z) f_k(x) \qquad (4.3.1)$$

Substituting (4.3.1) into the wave equation (1.4.4), making use of the properties of the HG basis functions and of the Gaussian Quadrature Formula for the Hermite polynomial, a set of coupled ordinary differential equations for the expansion coefficients is derived. The resulting set is written here in matrix form

$$\frac{d^2\underline{B}}{dz^2} + \underline{\underline{Q}}\,\underline{B} = 0 \qquad (4.3.2)$$

with $\underline{B} = [B_k(z); k = 1,2,\ldots,N]$ the column vector containing the field expansion coefficients and the matrix $\underline{\underline{Q}}$ is defined in equation (A4.3 - 6) of Appendix 4.3.

It is convenient to modify equation (4.3.2) in order to have the field $\underline{E}_{(M)}(z)$ as the unknown instead of the vector $\underline{B}(z)$. From equation (4.3.1) the following expression may be derived

$$\underline{E}_{(M)}(z) = \underline{\underline{G}}\,\underline{B}(z) \qquad (4.3.3)$$

with the matrix $\underline{\underline{G}}$ defined in Appendix 4.3, and hence equation (4.3.2) becomes

$$\frac{d^2\underline{E}(z)}{dz^2} + \underline{\underline{P}}(z)\underline{E}(z) = 0 \qquad (4.3.4)$$

with $\underline{\underline{P}}$ as defined in Appendix 4.3.

4. The collocation method

Within the category of methods of weighted residuals, such as the Galerkin method, the collocation method uses Dirac delta distributions as the weighting functions, [refer to Appendix 4.4], [7], [8], [9]. By combining the HG expansion method with the numerical collocation method (HGCM) the partial differential scalar wave equation (1.4.4) may be conveniently reduced to a set of coupled ordinary differential equations in the single variable, z.

Consider the general case of the field expansion (4.2.8) in a dielectric structure. In order to (numerically) solve equation (1.4.4) it is inevitable that the expansion (4.2.8) be limited to a finite number (M) of expansion terms

$$E_{(M)}(x,z) = \sum_{k=0}^{M} B_k(z) f_k(x) \qquad [4.3.1]$$

The next step is to discretise the transverse (x) axis into a set of points, x_i, called collocation points. The choice of collocation points is in principle arbitrary, but the set of the M zeros of the M-th order Hermite polynomial has been shown to be an

appropriate and convenient one [refer also to the use of the Gaussian Quadrature formula, as discussed in the previous Section 3].

For the present case, therefore, the collocation points are defined by

$$x_i: \; H_M(x_i) = 0 \quad i = 1,2,...,M \tag{4.4.1}$$

The field $E_{(M)}(x, z)$ can be sampled at the collocation points to obtain

$$E_{(M)}(x_i, z) = \sum_{k=0}^{M} B_k(z) f_k(x_i) \tag{4.4.2}$$

Equation (4.4.2) can be interpreted as the i-th component of the column vector containing the values of the field sampled at the collocation points:

$$E_i(z) = E_{(M)}(x_i, z) \tag{4.4.3}$$

and hence $\underline{E}_{(M)}(z) = [E_i(z); i = 1, 2, ..., M]$.

Compared to (4.2.8), the finite expansion (4.3.1) is not an exact solution of (1.4.4). The difference between the exact solution (4.2.8) and the approximate one (4.3.1) is estimated by the residual function R(x, z, M). For the particular case of the scalar wave equation (1.4.4) R(x, z, M) is defined by

$$R(x,z,M) = \left| \left[\nabla^2 - k_o^2 \varepsilon(x,z) \right] E_{(M)}(x,z) \right| \tag{4.4.4}$$

For the collocation method, the residual function R(x, z, M) is forced to vanish at the chosen collocation points (x_j) [Appendix 4.4]. This is equivalent to writing the scalar wave equation at each collocation point.

Thus, using equation (4.4.4) and the properties of the HG functions, the scalar wave equation (1.4.4) can be reduced to a matrix equation for the field $\underline{E}_{(M)}(z)$:

$$\frac{d^2 \underline{E}_{(M)}(z)}{dz^2} + \underline{\underline{S}}(z) \underline{E}_{(M)}(z) = 0 \tag{4.4.5}$$

where $\underline{\underline{S}}(z)$ is a (real) matrix which depends on z via the refractive index distribution, as presented in Appendix 4.5. It is worth noticing that equation (4.4.5) is equivalent to (4.3.4), thus confirming the correspondence between the collocation method and the analytic derivation discussed in the previous section.

In summary, the advantages of combining the collocation method with the HG expansion method are that i) the partial differential Helmholtz equation (1.4.4) can be transformed into a set of coupled ordinary differential equations; ii) no numerical integration is required; and iii) the analysis follows from a matrix formalism.

5. Choice of the width parameter (w_o) of the HG functions

When the field expansion is done with the complete (infinite) set of HG functions, as in (4.2.1) or (4.2.8) or (4.2.11), then any value for the width parameter, w_o, should yield the correct representation of the original function. For numerical computation, however, the number of expansion terms is finite and then the choice of w_o may affect the accuracy of the representation.

In particular situations, e.g., in the representation of a single mode of a three-layer symmetric slab (open) waveguide by an individual Hermite-Gauss function of the corresponding order, it is possible to use the variational method [Chapter 2, Section 3] to derive an optimum waist parameter w_o which gives the most accurate value for the propagation constant for that mode. With the variational method it is assumed that the (individual) HG function adequately represents the mode profile since the emphasis is on obtaining an accurate value only for the propagation constant. For multimoded waveguides each mode is considered separately and the corresponding individual width parameters do not necessarily have the same value. This highlights the problem that arises when applying the HGCM to eigenvalue problems related to multimoded waveguides, where all modes are computed simultaneously. The choice of w_o in the context of the numerical solution of field problems is presented and discussed in the next chapter.

For the solution of field propagation (initial value problems) in longitudinally non-uniform structures, it is the total field that is to be approximated by an expansion in terms of a finite number of HG functions, equation (4.3.1). The optimal choice of w_o in this case is the one that best approximates the field shape. Also in the case of initial value problems it is strictly not possible to determine w_o *a priori*, although a workable choice can be made in this respect. One of the factors to be taken into account in the selection of w_o is the transverse domain that has to be used for the computation (computational window). The collocation points, in fact, determine the domain on the transverse axis. Hence, as with all numerical procedures, there is a trade-off between the number of HG functions used and the extent of the lateral physical dimension that is included in the computations. [Note that the number M of collocation points is chosen to be the same as the number of terms used in the field expansion.]

The modelling of longitudinally non-uniform devices, e.g., flared geometry devices, may add further complications in the choice of w_o. The end facets may have significantly different widths, as for example shown in Fig. 4.5.1. The same value of w_o used all along the length of this device, for example, is unlikely to give satisfactory results. Analyses of this type of devices have shown that the use of appropriately different w_o parameters along the length increases the accuracy of the results. This has been achieved in the model presented in this thesis by using longitudinally piecewise constant w_o. A schematic illustrating this process is shown in Fig. 4.5.2.

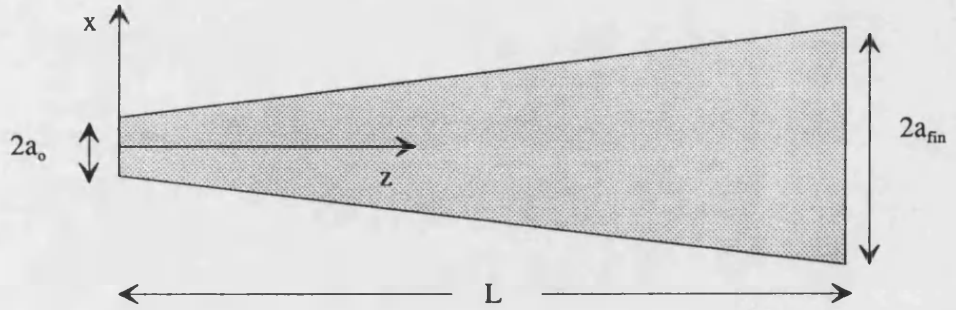


Fig. 4.5.1: Top view of a taper structure device: $2a_o$ and $2a_{fin}$ are the narrower and wider width, respectively, L is the length of the device.

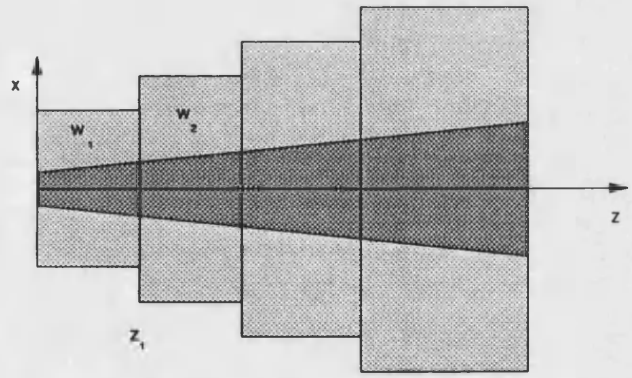


Fig. 4.5.2: Longitudinal sections, characterised by different values for the width parameter, e.g., w_1 and w_2 etc., for the analysis of tapers which have significantly varying widths.

Summary

The formalism for the electromagnetic field expansion using a set of orthogonal basis functions is presented in this chapter. In the present context it is argued that the HG functions are most apt since i) they form a complete set in the functional space of

interest, ii) they individually satisfy the radiation condition pertinent to optical structures, and iii) they form a discrete set.

Both the eigenvalue problem and the initial value problem have been analysed with the HG expansion scheme. Further, it is shown that the combination of the HG expansion with the collocation method produces a convenient scheme for field analysis in semiconductor optical structures. Finally, it is pointed out that no unique, optimum choice for w_o (the HG width parameter) is readily predictable for use with the HGCM even for the solution of the eigenvalue problem.

Appendix 4.1

HG expansion method for the eigenvalue problem

Consider the field expansion

$$F_m(x) = \sum_{k=0}^{\infty} a_k^{(m)} f_k(x) \qquad [4.2.1]$$

for the m-th mode of a (longitudinally uniform) waveguide, where $a_k^{(m)}$ are the (constant) expansion coefficients and $f_k(x)$ are the HG functions (for the sake of simplicity the width parameter is taken to be $w_o = 1$). Substitute (4.2.1) into the eigenvalue equation (1.10.3)

$$\frac{d^2F_m(x)}{dx^2} + \left(k_o^2\varepsilon - \beta_m^2\right)F_m(x) = 0 \qquad [1.10.3]$$

where $F_m(x)$, the modal field profile (eigenfunction), and β_m, the corresponding propagation constant (eigenvalue), are both still unknown. The resulting equation is

$$\sum_k a_k^{(m)}\left[f_k''(x) + \left(k_o^2\varepsilon - \beta_m^2\right)f_k(x)\right] = 0 \qquad (A4.1 - 1)$$

where $''$ depicts the second derivative with respect to x. Multiplying by $f^*_m(x)$, integrating over the infinite range $(-\infty,+\infty)$, and making use of

$$\left[\frac{d^2}{dx^2} + (-x^2 + 2m + 1)\right]f_k(x) = 0 \qquad [2.2.6]$$

obtain

$$\sum_k a_k^{(m)}\left[\int_{-\infty}^{+\infty} f_m^*(x^2 - 2k - 1)f_k\,dx + k_o^2\int_{-\infty}^{+\infty} f_m^*\varepsilon(x)f_k\,dx - \beta_m^2\int_{-\infty}^{+\infty} f_m^* f_k\,dx\right] = 0 \quad (A4.1 - 2)$$

with k = 1, 2, ... Consider the two cases of k = m and k ≠ m separately, where m is the waveguide mode of interest.

1. k = m

Assuming $a_m^{(m)} \neq 0$, equation (A4.1 - 2) becomes

$$\beta_m^2 + 2m + 1 - k_o^2\int_{-\infty}^{+\infty} f_m^*\left[\varepsilon(x) - \left(-\frac{x^2}{k_o^2}\right)\right]f_m\,dx = 0 \qquad (A4.1 - 3)$$

It is worth noting that equation (2.2.6) may be conveniently rewritten as

$$\frac{d^2 f_k(x)}{dx^2} + \left(k_o^2 \tilde{\varepsilon}(x) - \tilde{\beta}_k^2\right) f_k(x) = 0 \qquad \text{(A4.1 - 4)}$$

with $\tilde{\varepsilon}(x) = -\frac{x^2}{k_o^2} + b$ and $\tilde{\beta}_k^2 = k_o^2 b - (2k+1)$. Equation (A4.1 - 4) can be then interpreted as the equation of the unperturbed quadratic medium whose eigensolutions are the HG functions and the eigenvalues defined as above by $\tilde{\beta}_k$. In this context, then, equation (A4.1 - 3) (with k = m) becomes

$$\beta_m^2 - \tilde{\beta}_m^2 = k_o^2 \int_{-\infty}^{+\infty} f_m^*(x)[\varepsilon(x) - \tilde{\varepsilon}(x)] f_m(x) dx \qquad \text{(A4.1 - 5)}$$

which is equivalent to that derived with the perturbation method [Chapter 3, Section 1.1]. Summarising, for the case k = m, ε(x) in (1.10.3) is considered as a perturbation of $\tilde{\varepsilon}(x)$ (which is the unperturbed parabolic dielectric distribution).

2. k ≠ m

Equation (A4.1 - 2) becomes

$$\sum_k a_k^{(m)} \int_{-\infty}^{+\infty} f_m^*(x)\left[k_o^2 \varepsilon(x) + x^2\right] f_k(x) dx = 0 \qquad \text{(A4.1 - 6)}$$

the solution of which gives the expansion coefficients $a_k^{(m)}$ with k ≠ m (k = 1, 2, ...). The only undetermined coefficient $a_m^{(m)}$ is found by normalising the eigenfunction, which leads to

$$a_m^{(m)} = \sqrt{1 - \sum_{k \neq m} \left(a_k^{(m)}\right)^2} \qquad \text{(A4.1 - 7)}$$

for the m-th mode. The waveguide mode of order m can be thus reconstructed as the linear combination of HG functions (4.2.1), using the expansion coefficients specified by (A4.1 - 6) - (A4.1 - 7); the corresponding propagation constant is given by equation (A4.1 - 5).

Appendix 4.2

HG expansion method for the initial value problem

Consider the field expansion

$$E(x,z) = \sum_{k=0}^{\infty} B_k(z) f_k(x) \qquad [4.2.8]$$

with $f_k(x)$ the HG basis functions, (2.2.4), and $B_k(z)$ the (complex) expansion coefficients. Substituting (4.2.8) into the scalar wave equation (1.4.4), obtain

$$\sum_k \left\{ \frac{d^2 B_k(z)}{dz^2} f_k(\xi) + B_k(z) \left[\frac{1}{w_o^2} f_k''(\xi) + k_o^2 \varepsilon(\xi, z) f_k(\xi) \right] \right\} = 0 \qquad \text{(A4.2 - 1)}$$

where $\xi = \frac{x}{w_o}$ and $f_k''(\xi) = \frac{d^2 f_k(\xi)}{d\xi^2}$.

Making use of equation (2.2.6), (A4.2 - 1) becomes

$$\sum_k \left\{ \frac{d^2 B_k(z)}{dz^2} + B_k(z) \left[k_o^2 \varepsilon(\xi, z) + \frac{1}{w_o^2} (\xi^2 - 2k - 1) \right] \right\} f_k(\xi) = 0 \qquad \text{(A4.2 - 2)}$$

Multiplying by $f^*_j(x)$ and integrating over $(-\infty,+\infty)$ obtain

$$\sum_k \left\{ \frac{d^2 B_k(z)}{dz^2} \int_{-\infty}^{+\infty} f_j^*(\xi) f_k(\xi) d\xi + B_k(z) \int_{-\infty}^{+\infty} f_j^*(\xi) \left[\left(\frac{\xi}{w_o} \right)^2 - \frac{2k+1}{w_o^2} + k_o^2 \varepsilon(\xi, z) \right] f_k d\xi \right\} = 0 \qquad \text{(A4.2 - 3)}$$

and hence

$$\frac{d^2 B_j(z)}{dz^2} - \frac{2j+1}{w_o^2} B_j(z) + \sum_k B_k(z) \int_{-\infty}^{+\infty} f_j^*(\xi) \left[\left(\frac{\xi}{w_o} \right)^2 + k_o^2 \varepsilon(\xi, z) \right] f_k(\xi) d\xi = 0 \qquad \text{(A4.2 - 4)}$$

with j = 1, 2, ... In a compact form equation (A4.2 - 4) is written as

$$\frac{d^2 B_j(z)}{dz^2} - \frac{2j+1}{w_o^2} B_j(z) + \sum_k Q_{jk}(z) B_k(z) = 0 \qquad \text{(A4.2 - 5)}$$

where j = 1, 2, ... and

$$Q_{jk}(z) = \int_{-\infty}^{+\infty} f_j^*(\xi) \left[\left(\frac{\xi}{w_o} \right)^2 + k_o^2 \varepsilon(\xi, z) \right] f_k(\xi) d\xi \qquad \text{(A4.2 - 6)}$$

is the matrix element.

Appendix 4.3

Analytic method for the discretisation of the system of coupled differential equations (4.2.9) for the expansion coefficients $B_k(z)$ - derivation of equation (4.3.4)

The starting equation is (4.2.9) which has been derived by expanding the total field, E(x,z), as in (4.2.8). It is convenient to write the set of equations (4.2.9) in the following explicit form, using (2.2.4),

$$\frac{d^2B_k(z)}{dz^2} - \frac{2k+1}{w_o^2}B_k(z) + \sum_j B_j(z)C_kC_j\int_{-\infty}^{+\infty} H_k(\xi)\left[\left(\frac{\xi}{w_o}\right)^2 + k_o^2\varepsilon(\xi,z)\right]H_j(\xi)e^{-\xi^2}d\xi = 0 \qquad \text{(A4.3 - 1)}$$

where k = 1, 2, ... and $\xi = \frac{x}{w_o}$.

The first step is to discretise the integrals in (A4.3 - 1), which also implies the discretisation of the transverse axis. This is achieved by means of the Gaussian Quadrature formula for the Hermite polynomials, [6],

$$\int_{-\infty}^{+\infty} g(\xi)e^{-\xi^2}d\xi \approx \sum_{i=1}^{P} h_i g(\xi_i) \qquad \text{(A4.3 - 2)}$$

for any function $g(\xi)$, where P is the number of (sampling) points ξ_i on the transverse axis. The Gaussian Quadrature formula for Hermite polynomials imposes that the sampling points be the zeros of the P-th order Hermite polynomial. [The same set of points are shown to be a convenient choice for the collocation points]. The weight functions h_i are defined by

$$h_i = \frac{2^{P-1}P!\sqrt{\pi}}{P^2\left[H_{P-1}(\xi_i)\right]^2} \qquad \text{(A4.3 - 3)}$$

Both the set of sampling points and the weight functions are specified by the orthogonal polynomials used in the Gaussian Quadrature formula (which exists for different types of polynomials, [6]).

Using (A4.3 - 2) the set (A4.3 - 1) becomes

$$\frac{d^2B_k(z)}{dz^2} - \frac{2k+1}{w_o^2}B_k(z) + \sum_j B_j(z)C_kC_j\sum_{i=0}^{P} H_k(\xi_i)\left[\left(\frac{\xi_i}{w_o}\right)^2 + k_o^2\varepsilon(\xi_i,z)\right]h_iH_j(\xi_i) = 0 \qquad \text{(A4.3 - 4)}$$

with k = 1, 2, ... The formal version of equation (A4.3 - 4) is for the series expansion (4.2.8). However, in practice, only a finite number of terms, M, is used, so that (4.2.8) becomes

$$E_{(M)}(x,z) = \sum_{k=1}^{M} B_k(z) f_k(x) \qquad [4.3.1]$$

It is then convenient to choose M = P so that the number of expansion terms is the same as the number of sampling points. The set (A4.3 - 4) can thus be written in a compact form making use of the matrix formalism (square matrices of order M):

$$\frac{d^2\underline{\underline{B}}}{dz^2} + \underline{\underline{Q}}\,\underline{\underline{B}} = 0 \qquad (A4.3 - 5)$$

where $\underline{\underline{B}} = [B_k(z); k = 1,2,\ldots,M]$ is the column vector of the expansion coefficients at a specified position z;

$$\underline{\underline{Q}} = \underline{\underline{G}}^{-1}\underline{\underline{D}}_1\underline{\underline{G}} + \underline{\underline{G}}^{-1}\underline{\underline{R}}\,\underline{\underline{G}} - \underline{\underline{D}}_2 \qquad (A4.3 - 6)$$

with $\underline{\underline{G}} = \{f_k(\xi_i); i,k = 1,2,\ldots,M\}$ the matrix whose columns are the HG functions sampled at the sampling points ξ_i;

$\underline{\underline{D}}_1 = \frac{1}{w_o^2}\mathrm{diag}\{\xi_i^2, i = 1,2,\ldots,M\}$ and $\underline{\underline{D}}_2 = \frac{1}{w_o^2}\mathrm{diag}\{(2i-1), i = 1,2,\ldots,M\}$ two diagonal matrices;

$\underline{\underline{R}}(z) = k_o^2\,\mathrm{diag}\{n^2(\xi_i, z), i = 1,2,\ldots,M\}$ the diagonal matrix of the dielectric distribution sampled at ξ_i.

The discretisation of the transverse axis permits to write the total field function E(x, z) at each sampling point x_i as

$$E_{(M)}(x_i,z) = \sum_{k=1}^{M} B_k(z) f_k\left(\frac{x_i}{w_o}\right) \qquad (A4.3 - 7)$$

which can be regarded as the i-th component of the M-dimensional column vector $\underline{E}(z)$, i.e.,

$$E_i(z) = E(x_i,z) \qquad (A4.3 - 8)$$

where the subscript (M) has been suppressed.

Hence $\underline{E}(z) = [E_i(z), i = 1, 2, \ldots M]$ can be also expressed as

$$\underline{E}(z) = \underline{\underline{G}}\,\underline{B}(z) \qquad (A4.3 - 9)$$

Using equation (A4.3 -9) the matrix equation (A4.3 - 5) can be rewritten in terms of the (sampled) total field function:

$$\frac{d^2\underline{\underline{E}}(z)}{dz^2} + \underline{\underline{P}}(z)\underline{\underline{E}}(z) = 0 \qquad \text{(A4.3 - 10)}$$

where $\underline{\underline{P}} = \underline{\underline{D}}_1 + \underline{\underline{R}} - \underline{\underline{G}}\underline{\underline{D}}_2\underline{\underline{G}}^{-1}$.

Appendix 4.4

The residual function

In the method of weighted residuals the residual function is forced to be zero in an average sense; i.e., the weighted integrals of the residual are forced to vanish so that

$$\int_I w_i R(x, z, M) dx = 0 \qquad \text{(A4.4 - 1)}$$

w_i are the weighting functions, M is the discretisation parameter, and I is the range of integration. Different methods of weighted residuals can be found in the literature, each characterised by a particular choice of weighting w_i. For the collocation method the weighting functions are the Dirac delta distributions centered at the collocation points, i.e.,

$$w_i = \delta(x - x_i) \qquad \text{(A4.4 - 2)}$$

where i = 1, 2, ..., M, and x_i are the collocation points. The definition of weighting function given in (A4.4 - 2) is equivalent to forcing the residual function R(x, z, M) to vanish at the collocation points, i.e.,

$$R(x_i, z, M) = 0 \qquad \text{(A4.4 - 3)}$$

As expected, the accuracy of the method increases with M, [9].

Appendix 4.5

The collocation method - formalism

The approximate solution of the wave equation (1.4.4), $E_{(M)}(x, z)$, is expressed as a linear combination of M Hermite-Gauss functions, $f_m(x)$, as in equation (4.3.1). At each collocation point, x_i, $E_{(M)}(x_i, z)$ is a function only of z:

$$E_{(M)}(x_i, z) = E_{(M),i}(z) = \sum_{k=0}^{M} B_k(z) f_k(x_i) \qquad \text{(A4.5 - 1)}$$

From (A4.5 -1), $E_{(M)}(x_i, z)$ can be considered as the i-th component of the column vector $\underline{E}_{(M)}(z)$ which, then, represents the transverse distribution of the field at any longitudinal position z, sampled at the collocation points, x_i.

The residual function R(x,z,M) [Appendix 4.4] is made to vanish at the collocation points. This process is equivalent to writing Helmholtz equation, (1.4.4), at each collocation point, i.e.,

$$\partial_x^2 E_{(M)}(x,z)\Big|_{x=x_i} + \frac{d^2 E_{(M),i}(z)}{dz^2} + k_o^2 \varepsilon(x_i, z) E_{(M),i}(z) = 0 \qquad \text{(A4.5 - 2)}$$

where i = 1, 2, ... M; the subscript (M) will be henceforth suppressed. In a compact (matrix) form (A4.5 - 2) becomes

$$\frac{d^2 \underline{E}(z)}{dz^2} + \underline{D}(z) + \underline{\underline{R}}(z)\underline{E}(z) = 0 \qquad \text{(A4.5 - 3)}$$

where $$\underline{D}(z) = \left[\partial_x^2 E(x,z)\Big|_{x=x_i},\ i = 1,2,\ldots M\right] = \left[\sum_{k=0}^{M} B_k(z) \frac{d^2 f_k(x)}{dx^2}\Big|_{x=x_i},\ i = 1,2,\ldots M\right]$$

and $\underline{\underline{R}}(z)$ is the diagonal matrix of the dielectric distribution sampled at the collocation points, $\underline{\underline{R}}(z) = k_o^2 \mathrm{diag}\{\varepsilon(x_i, z), i = 1,2,\ldots M\}$.

Since the HG functions, $f_k(x)$, satisfy the Sturm-Liouville differential equation

$$\frac{d^2 f_k(x)}{dx^2} + (2k + 1 - x^2) f_k(x) = 0 \qquad \text{[2.2.6]}$$

the vector $\underline{D}(z)$ can be reduced to an expression involving only the vector $\underline{E}(z)$ as the unknown, as shown below.

$\underline{D}(z)$ can be written as

$$\underline{D} = \underline{\underline{\tilde{D}}}\underline{B}(z) \qquad \text{(A4.5 - 4)}$$

where $\underline{\underline{\tilde{D}}} = \left\{ \left. \frac{d^2 f_k(x)}{dx^2} \right|_{x=x_i} ; i,k = 1,2,\ldots M \right\}$ is the matrix of the second derivatives of the HG functions. Further, using (2.2.6), equation (A4.5 - 4) reduces to

$$\underline{\underline{\tilde{D}}} = \underline{\underline{D}}_1 \underline{\underline{G}} - \underline{\underline{G}}\underline{\underline{D}}_2 \qquad \text{(A4.5 - 5)}$$

with $\underline{\underline{G}} = \{f_k(x_i); i,k = 1,2,\ldots M\}$ the matrix whose columns are HG functions sampled at the collocation points, and $\underline{\underline{D}}_1 = \frac{1}{w_o^2} \text{diag}\left\{ \left(\frac{x_i^2}{w_o^2} \right), i = 1,2,\ldots M \right\}$ and $\underline{\underline{D}}_2 = \frac{1}{w_o^2} \text{diag}\{(2i-1),\ i = 1,2,\ldots M\}$ two diagonal matrices.

It is also convenient to write the field sampled at the collocation points in the form $\underline{E}(z) = \underline{\underline{G}}\underline{B}(z)$, so that

$$\underline{B}(z) = \underline{\underline{G}}^{-1}\underline{E}(z) \qquad \text{(A4.5 - 6)}$$

Using (A4.5 -6), equation (A4.5 - 4) becomes

$$\underline{D} = \underline{\underline{\tilde{D}}}\underline{B} = \left(\underline{\underline{D}}_1 \underline{\underline{G}} - \underline{\underline{G}}\underline{\underline{D}}_2\right)\underline{\underline{G}}^{-1}\underline{E} = \left(\underline{\underline{D}}_1 - \underline{\underline{G}}\underline{\underline{D}}_2\underline{\underline{G}}^{-1}\right)\underline{E} \qquad \text{(A4.5 -7)}$$

Thus, making use of (A4.5 - 5) and (A4.5 - 7) the matrix equation (A4.5 - 3) becomes

$$\frac{d^2\underline{E}(z)}{dz^2} + \underline{\underline{S}}(z)\underline{E}(z) = 0 \qquad \text{(A4.5 - 8)}$$

with $\underline{\underline{S}}(z) = \underline{\underline{R}}(z) + \underline{\underline{D}}_1 - \underline{\underline{G}}\underline{\underline{D}}_2\underline{\underline{G}}^{-1}$ a real matrix which depends on z via the matrix of the dielectric distribution $\underline{\underline{R}}(z)$ only.

1. Inversion of matrix $\underline{\underline{G}}$

The matrix $\underline{\underline{G}}$ always has the inverse, $\underline{\underline{G}}^{-1}$, because the HG functions are orthogonal and the collocation points are all distinct. Making use of the properties of the HG functions and of the Gaussian Quadrature formula it is possible to determine the inverse of matrix $\underline{\underline{G}}$ algebraically. Consider the orthogonality property

$$\int_{-\infty}^{+\infty} f_j^*(x) f_k(x) dx = \delta_{jk} \qquad \text{[2.2.7]}$$

and the Gaussian Quadrature Formula applied to the same pair of functions

$$\int_{-\infty}^{+\infty} f_j^*(x) f_k(x) dx \approx \sum_{i=1}^{M} f_j^*(x_i) h_i e^{+x_i^2} f_k(x_i) \qquad \text{(A4.5 - 9)}$$

with the weight function h_i as defined in Appendix 4.3, equation (A4.3 - 3). By comparison of (2.2.7) with (A4.5 -9) obtain the following matrix equation

$$\underline{\underline{G}}^T \underline{\underline{H}}\underline{\underline{G}} = \underline{\underline{1}} \qquad \text{(A4.5 - 10)}$$

where T depicts the transpose matrix, $\underline{\underline{H}} = \text{diag}\left\{h_i e^{x_i^2}; i = 1,2,\ldots,M\right\}$, and $\underline{\underline{1}}$ is the unity matrix. Thus, from (A4.5 - 10) it follows that

$$\underline{\underline{G}}^{-1} = \underline{\underline{G}}^T \underline{\underline{H}} \qquad \text{(A4.5 - 11)}$$

References

[1] D. C. Stinson, *Intermediate mathematics of electromagnetics*, Prentice-Hall Electrical Engineering Series (1976)

[2] E. Butkov, *Mathematical Physics*, Addison-Wesley Publishing Company (1973)

[3] D. Marcuse, *Light transmission optics*, Van Nostrand Reinholds (1982)

[4] E. K. Sharma, M. P. Singh, A. Sharma, *Variational analysis of optical fibers with loss or gain*, Optics Letters, 18 (1993), pp. 2096-2098

[5] A. Kostenbauer, Y. Sun, A. E. Siegman, *Eigenmode expansions using biorthogonal functions: complex-valued Hermite-Gaussians*, Journal of the Optical Society of America A, 14, n. 8 (1997), pp. 1780-1790

[6] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications (1965)

[7] R. F. Harrington, *Field computation by moment methods*, Malabar Krieger (1968)

[8] L. Collatz, *The numerical treatment of differential equations*, Springer-Verlag (1966)

[9] B. A. Finlayson, *The method of weighted residuals and variational principles (with applications in fluid mechanscs, heat and mass transfer)*, Academic Press (1972)

Chapter 5

Field analysis - results

In this chapter the Hermite-Gauss Collocation Method (HGCM), [Chapter 4, Section 4] is applied to the analysis of passive structures and in most cases it has been possible to compare the results obtained with the HGCM with those from other methods of solution.

In the first part of the chapter, Section 1, important aspects concerning the choice of w_o for the HG basis functions are discussed to underline the caution that must be exercised in the numerical solution of field problems. It is also demonstrated that the HGCM numerical results are remarkably improved by the use of a continuously varying (analytic) refractive index distribution (compared to those obtained with an abrupt, step index distribution). The HGCM is used to solve the eigenvalue problem related to different types of waveguides - including waveguides with loss/gain introduced through the imaginary part of the refractive index.

In the second part attention is focused on the computation of electromagnetic field propagation which is solved as an initial value problems. Free-space propagation is considered first, in Section 2.1. It is well known that diffraction problems can be conveniently solved using the Hermite-Gauss Beam analysis [Chapter 2, Section 1]. However, with the results shown here it is argued that the HG expansion scheme provides an alternative and, at least, an equally effective method. Problems of field propagation in (longitudinally uniform) waveguides (Section 2.2) and coupled waveguides (Section 2.3) are thus solved with the HGCM as test cases to assess the accuracy of the HGCM as a scheme for field propagation analysis.

The analysis of step discontinuities is presented in Section 2.4 to demonstrate that field analysis by means of the HGCM implicitly takes into account the contribution of the radiation modes. A second approximate method of solution based on a combination of bound modes and plane wave expansion is also presented in Section 2.4 to provide for another method against which the results from the HGCM are compared.

Tapered structures are next investigated: the HGCM is used to solve for field propagation in linear (Section 2.5) and parabolic (Section 2.6) shaped tapers. The

results obtained with the HGCM are compared with those obtained with an approximated mode-matching method, [1], [2]. Finally, in Section 2.7 some aspects of the numerical analysis of the HGCM are discussed in the context of the initial value problem.

1. Eigenvalue problem

1.1 Role of the w_o parameter in the HG function expansion

The Hermite-Gauss functions form a complete orthogonal set in the space of square integrable functions. Any function in the functional space of interest can thus be represented in terms of the HG set. Hence, for example, for a waveguide (bound) mode, $F_m(x)$, which is a square integrable function, it is possible to write

$$F_m(x) = \sum_{k=0}^{\infty} a_k^{(m)} f_k\left(\frac{x}{w_o}\right) \tag{5.1.1}$$

where w_o is the waist parameter of the HG basis functions, $f_k\left(\frac{x}{w_o}\right)$. The important point in equation (5.1.1) is that any value for the waist parameter, w_o, can be used with the HG set if the whole (infinite) set is considered in the series expansion.

For practical solutions, however, a finite number of expansion terms are used in (5.1.1), and hence

$$F_m(x) \approx \sum_{k=0}^{M} a_k^{(m)} f_k\left(\frac{x}{w_o}\right) = F_{m,(M)}(x) \tag{5.1.2}$$

The error in the approximation used in (5.1.2) is given by

$$\Delta F = F_m(x) - F_{m,(M)}(x) = \sum_{k=M+1}^{\infty} a_k^{(m)} f_k\left(\frac{x}{w_o}\right) \tag{5.1.3}$$

and depends on the waist parameter, w_o, and on the number of expansion terms, M, used in (5.1.2). With the finite expansion (5.1.2) it is thus important to choose w_o in such a way that ΔF is minimised. This problem begins to become significant only for relatively small values of M.

A particular case of (5.1.2) is when M = 1, which implies that a single HG function is used to approximate the field profile. The variational method has been used to determine the waist parameter of the best fitting HG function for that particular

mode, by minimising the error in the evaluated modal propagation constant, [refer to Chapter 2, Section 3].

For the most part in this thesis M = 99 is used in equation (5.1.2) with the HGCM, which is sufficiently large so that, in practice any reasonable w_o yields sufficiently accurate results. In this thesis the HG expansion method has been combined with the numerical collocation method, since the latter is a most convenient process with which to carry out the computation required for the HG expansion method.

However for the solution of step index profile waveguides numerical inaccuracies have been encountered. In fact it is noted that the collocation method is a sampling technique, and hence a step function is in general randomly sampled, depending on the position of the collocation points with varying w_o, and it is this that produces numerical inaccuracies rather than the finite number of expansion terms M used in (5.1.2).

The latter problem has been overcome by introducing a novel approach which uses a sharp, but continuously varying analytic (supergaussian) index distribution to approximate the step index profile. With this innovation the improvement in the computed results has been remarkable, and it is found that, as expected, the results are now relatively insensitive in w_o. Results confirming the conclusions drawn in the present section are discussed in the following subsections.

1.2 Piecewise constant refractive index distribution

For HGCM solutions of the eigenvalue problem for step index (piecewise constant) profile waveguides, the choice of w_o for the HG functions is important. As a general feature of the HGCM it is found that by varying w_o the value of the computed effective refractive index of a particular (bound) mode fluctuates around the correct value. This effect is shown in Fig. 5.1.1, where the eigenvalues for a three-layer, symmetric, multimoded slab waveguide calculated with the HGCM are plotted against the w_o parameter used in the calculations. The multimoded waveguide chosen for this example supports two bound modes, characterised by the effective refractive indices $n_{eff(o)}$ and $n_{eff(1)}$, respectively, and hence the two curves shown in the graph are

obtained simultaneously with the HGCM. The effective refractive indices determined analytically are also shown.

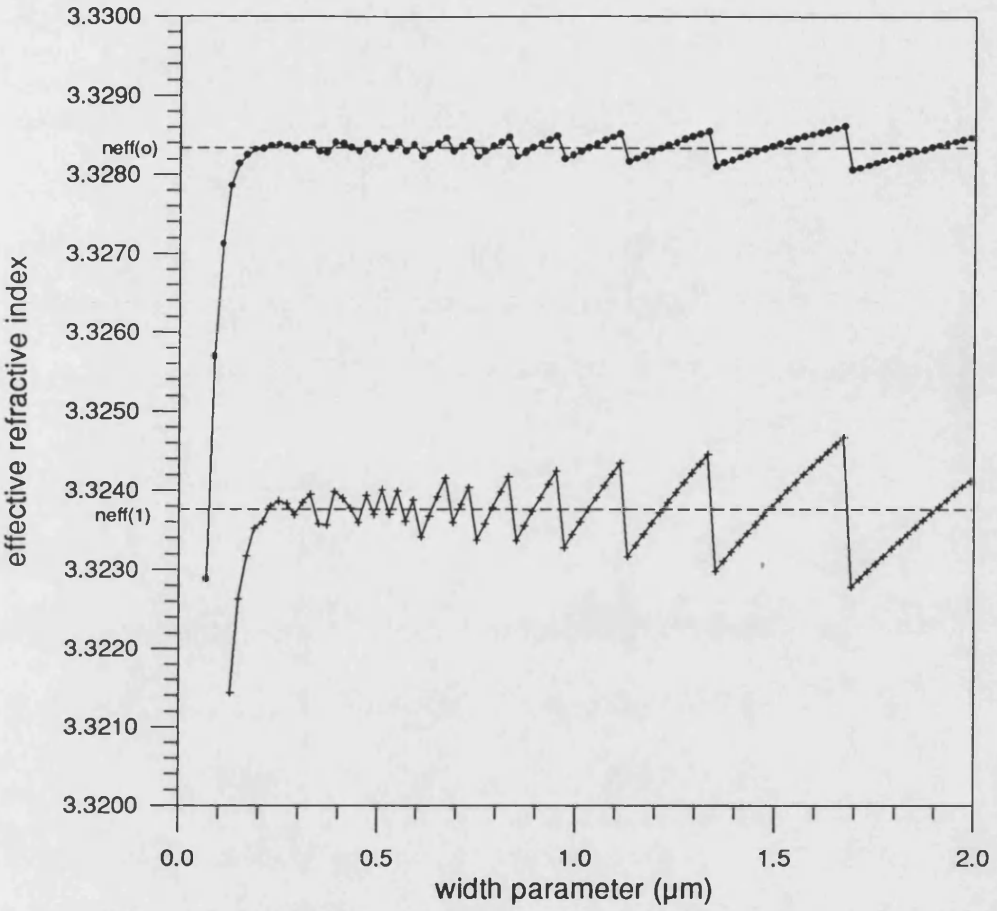


Fig. 5.1.1: Effective refractive index as a function of the width parameter, w_o, used in the HGCM calculations. The multimoded waveguide is characterised by the following parameters: $2a_o = 3\mu m$, $\lambda_o = 0.86\mu m$, $n_1 = 3.33$, $n_2 = 3.32$, $n_{eff(o)} = 3.32834$, $n_{eff(1)} = 3.32376$.

1.3 Continuously varying refractive index profile

The numerical inaccuracies discussed in the above Section 1.2 are dramatically reduced if the piecewise constant refractive index distribution (ST) is replaced by a continuously varying (analytically expressible) refractive index distribution. An abrupt change in index profile may be conveniently represented by the supergaussian (SG) analytic distribution

$$n(x) = n_2 + (n_1 - n_2)\exp\left[-\frac{1}{2}\left(\frac{x}{a}\right)^q\right] \qquad (5.1.4)$$

where $q \geq 10$ gives a supergaussian flat-top profile and 2a is the width of the waveguide. An example of supergaussian distribution (q = 16) is shown in Fig. 5.1.2, and compared with the piecewise constant distribution. In most cases step changes in index are not realised in practice, and so the supergaussian, in fact, represents a realistic index distribution.

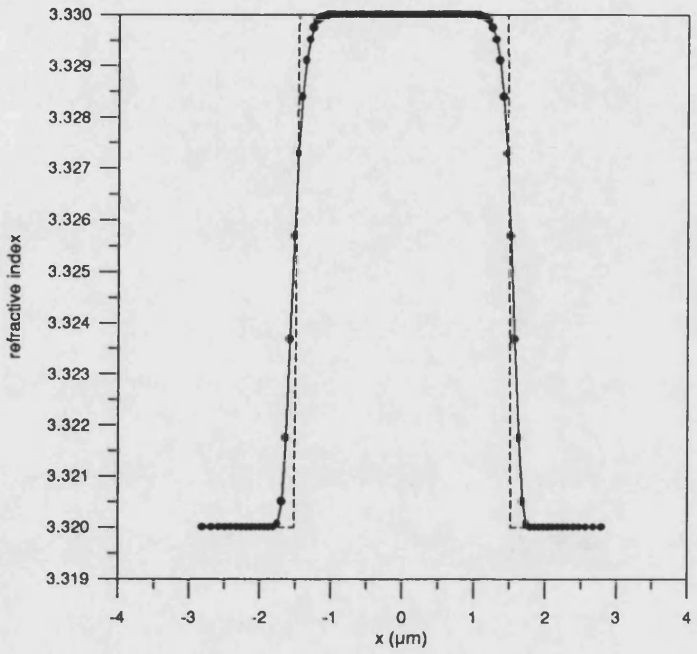


Fig. 5.1.2: Refractive index distribution (n_1 = 3.33, n_2 = 3.32, $2a_o$ = 3μm): solid line: supergaussian (q = 16); broken line: piecewise constant.

As discussed in Section 1.1 of this Chapter, it is preferable to use a continuous dielectric function with the HGCM because the collocation method is implicitly a sampling technique - this aspect of the HGCM is also mentioned in [3], [4] - and hence the accuracy of the results calculated with the HGCM increases significantly when the function to be sampled is analytic. The improvement is clearly evident by comparing Fig. 5.1.1 with Fig. 5.1.3.

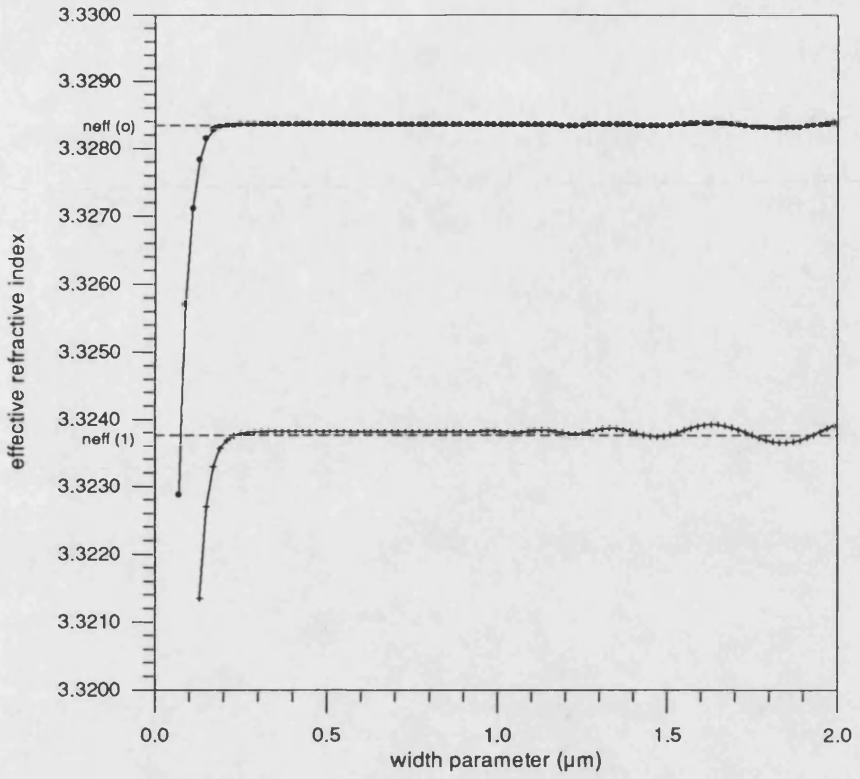
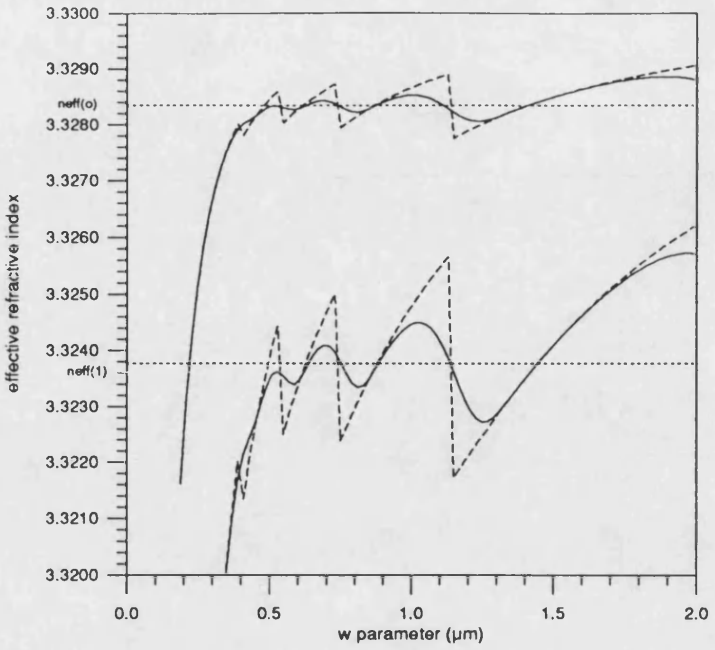


Fig. 5.1.3: Effective refractive index as a function of the w_o parameter used in the HGCM calculations using a continuous SG refractive index distribution (q = 16). (Parameters in Fig. 5.1.1)

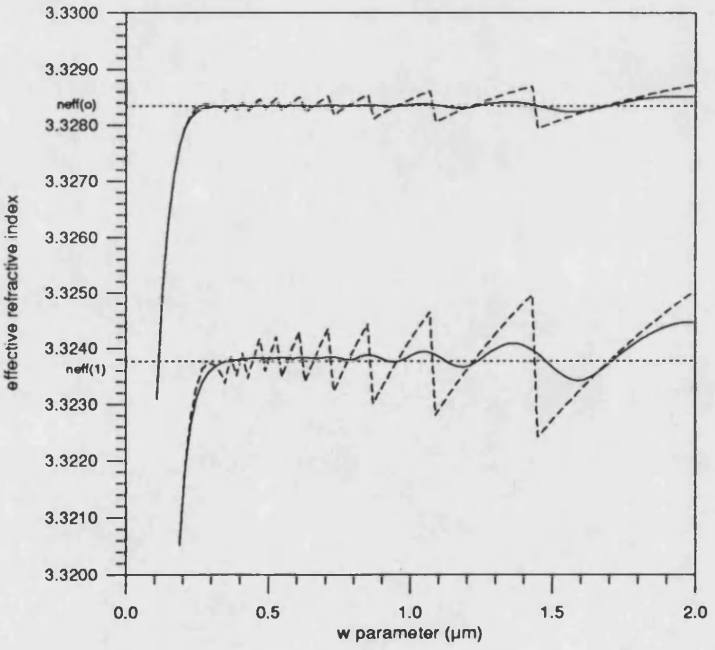
1.4 Number of expansion terms, M

The other parameter to be taken into account in the HGCM is the number, M, of expansion terms used in the field expansion (5.1.1), which corresponds to the number of collocation points. The effect of M on the (computed) effective refractive index of the bound modes of the eigenvalue problem (3.1.1) is shown in Fig. 5.1.4. From the discussions of the preceding Section 1.1 it is clear that by increasing the number of expansion terms the approximation (5.1.1) approaches the limit case of the series expansion. Consequently, it is seen that the number of w_o values that give the correct eigenvalues increases with M. However, it emerges again that the HGCM is far more efficient when a continuous SG refractive index profile is used in the calculations although, even for the ST refractive index profile the improvement is evident upon increasing M. Finally, it is worth mentioning that for small M values it has been noticed that spurious solutions may occur when using large values for w_o for the solution of the ST refractive index distribution, which could be attributed to the

incorrect definition of the width of the waveguide rib and to the poor accuracy associated with low values of M.



a)



b)

Fig. 5.1.4: Effective refractive index as a function of the w_o parameter and of the number of expansion term used in the HGCM calculations: a) M = 11, b) M = 41. The curves refer to broken line ST refractive index distribution, solid line SG refractive index distribution (q = 16). (Parameters in Fig. 5.1.1)

1.5 Waveguide bound modes - real refractive index distribution

The HGCM has been tested solving for the bound modes of slab waveguides defined by several (real) refractive index distributions, n(x). These results are compared with those from the Transfer Matrix Method (TMM) [5], and, whenever possible, with the analytic solution (A). The results are shown in Tab. 5.1.1. The values of the effective refractive index of the (bound) modes computed with the HGCM are in good agreement with those calculated with the other two methods. The relative error is defined by

$$\delta = \left| \frac{n_{eff\,(HGCM)} - n_{eff\,(OTHER\ METHOD)}}{n_1 - n_2} \right| \times 100 \qquad (5.1.5)$$

and is of the order of 0.1%. Very good agreement is also found for the corresponding modal profiles.

Note that in Tab. 5.1.1 the last set of results refer to the TE/TM mode propagation constants of single mode waveguides characterised by a Gaussian dielectric distribution. In this case the HGCM has been used to solve the two different scalar equations that generate TE and TM mode solutions, as discussed in Appendix 1.4. It is satisfying to observe that the two (different) propagation constants could be determined by the HGCM with sufficient accuracy.

Table 5.1.1: Effective refractive index for waveguide bound modes calculated with HGCM, TMM and analytic solution (A)

λ (μm)	n_1	n_2	$2a_o$ (μm)	M	w_o (μm)	n_{eff} (HGCM)	n_{eff} (TMM)	n_{eff} (A)
0.75	3.5	3.35	0.1	61	0.117	3.3723	3.3723	3.3723
1.0	3.5	3.3541	0.1	61	0.120	3.3672	3.3671	3.3671
0.75	3.5	3.4641	0.1	61	0.209	3.4656	3.4655	3.4655
1.5	3.5	3.35	0.4	99	0.106	3.4088	3.4082	3.4082
1.0	3.5	3.35	0.1	61	0.091	3.3637	3.3637	3.3637
(+) 0.75	3.5	3.304	0.1	99	0.045	3.3196	3.3196	
(*) 1.0	3.5	3.35	0.2	99	0.08	3.3772	3.3787	
(^) 1.55	3.42	3.4187	5.0	31	2.5	3.419207	3.419206	
(") 0.86	3.33	3.32	4.0	99	1.051	I 3.3289	3.3289	3.3289
						II 3.3258	3.3258	3.3266
						III 3.3213	3.3213	3.3213
(T) 1.55	3.42	3.4187	4.0	99	1.95	TE 3.4190424	TE 3.4190427	
						TM 3.4190418	TM 3.4190426	

n_1, n_2 = highest and lowest values of the refractive index; unless otherwise specified, the waveguide structure is a three-layer slab

$2a_o$ = thickness of the waveguiding layer

n_{eff} = effective refractive index of the mode

M = number of collocation points

w_o = width parameter of the HG functions

(+) truncated ***parabolic*** refractive index distribution: $n(x) = \begin{cases} n_1 - (n_1 - n_2)\left(\frac{x}{a_o}\right)^2 & |x| \le a_o \\ n_2 & |x| \ge a_o \end{cases}$

(*) ***Gaussian*** refractive index profile: $n^2(x) = n_2^2 + (n_1^2 - n_2^2)e^{-\frac{x^2}{a_o^2}}$

(^) ***asymmetric*** refractive index profile: $n(x) = \begin{cases} n_1 & x < a_o \\ n_2 & |x| \le a_o \\ n_3 & x > a_o \end{cases}$ with $n_3 = 3.4183$

(") multimoded waveguide: I, II, III refer to the fundamental, first and second order mode, respectively

(T) singlemoded waveguide characterised by a Gaussian profile, as in (*): TE and TM refer to the propagation constants of the TE and TM modes, respectively

Another advantage of using the HGCM is that the propagation constants and transverse profiles of all the bound modes supported by the waveguide are calculated simultaneously by solving the eigenvalue matrix equation (3.1.1).

In conclusion to this section note that it is possible to redistribute the collocation points by means of a transformation of variables, as described in [4], [6]. The resulting modified HGCM may be convenient in some specific cases, e.g., in the calculation of the eigenvalues associated with a (longitudinally uniform) waveguide. Making use of the transformation of variables the density of points in the central part of the transverse domain (core region of the waveguide) may be increased. As a consequence the points in the outer region (decaying part of the field) are reduced in number, but at the same time they are spread out and hence cover a wider transverse range. This procedure, described in Appendix 5.1, is particularly useful in the analysis of strong waveguides because the field is tightly confined in the core region. For weaker waveguides, the field spreads out in the cladding layers, so that an almost uniformly distributed set of collocation points (such as the set of zeros of Hermite polynomials) is desirable.

1.6 Waveguide bound modes - complex refractive index distribution

The Hermite-Gauss Collocation Method has also been applied to solve for the complex eigenvalue problem that corresponds to the modes of waveguides with complex refractive index distribution, i.e., waveguides with gain or loss, [Chapter 4, Section 2.1.1]. Specifically, three-layer symmetric slab waveguides with complex refractive index in the core region, $\hat{n}_1 = n_{1,R} + in_{1,I}$, are considered.

The results are summarised in Tab. 5.1.2 where the results calculated with the HGCM are compared with the numerical solution of the complex dispersion relation (D). The agreement between the two sets of results is satisfactory. The HGCM provides the values for the real part of the mode effective refractive index within a relative error, defined by

$$\delta_R = \left| \frac{n_{eff,R}^{(HGCM)} - n_{eff,R}^{(A)}}{n_{1,R} - n_{2,R}} \right| \times 100 \qquad (5.1.6)$$

of the order of 0.1 %. Also the results for the imaginary part of the eigenvalues calculated with the HGCM are sufficiently accurate, the discrepancy, defined by

$$\delta_I = \left| \frac{n_{eff,I}^{(HGCM)} - n_{eff,I}^{(A)}}{n_{1,I} - n_{2,I}} \right| \times 100 \qquad (5.1.7)$$

being of the order of 10%.

Table 5.1.2: Effective refractive index of the (bound) modes of three-layer slab waveguides with complex dielectric distribution in the core layer.

	w (μm)	n_{eff} (HGCM, M=99)	n_{eff} (D)
1	0.105	(3.4655, 7.4 10^{-5})	(3.4655, 4.5 10^{-5})
2	0.595	(3.3448, 6.9 10^{-5})	(3.3448, 7.1 10^{-5})
(*) 3	2.0	(3.4999, 2.3 10^{-6})	(3.4999, 2.8 10^{-6})
(+) 4	1.048	I (3.3289, 9.7 10^{-5})	I (3.3289, 9.7 10^{-5})
		II (3.3258, 9.0 10^{-5})	II (3.3258, 1.1 10^{-5})
		III (3.3213, 6.5 10^{-5})	III (3.3213, 6.3 10^{-5})
(+) 5	1.61	I (1.8729, 6.0 10^{-4})	I (1.8729, 6.19 10^{-4})
		II (1.8670, 5.8 10^{-4})	II (1.8699, 6.7 10^{-4})
		III (1.8577, 5.0 10^{-4})	III (1.8578, 5.1 10^{-4})

where

	λ (μm)	2a (μm)	n_1	n_2
1	0.75	0.1	(3.5, 5.5724 10^{-4})	3.4641
2	1.5	2.0	(3.35, 1.0 10^{-4})	3.34
(*) 3	0.86	4.0	(3.5, 1.0 10^{-4})	3.5
(+) 4	0.86	4.0	(3.33, 1.0 10^{-4})	3.32
(+) 5	0.85	4.0	(1.875, 6.28 10^{-4})	1.85

(*) note that in this particular example there is no real refractive index step

(+)multimoded waveguide: I, II, III refer to the fundamental, first and second order mode, respectively

2. Initial value problem

2.1 Free-space propagation

Free-space propagation provides a quite challenging example for testing the field propagation scheme based on the Hermite-Gauss expansion method. This diffraction problem can be solved analytically in the paraxial approximation (Fresnel diffraction). The well-known analytic solutions are the Hermite-Gauss Beams, [Chapter 2, Section 1]. Hence, given a Gaussian initial field distribution, the field diffracts in the homogeneous medium maintaining the Gaussian profile, but increasing in (lateral) dimension because of the longitudinally varying beam waist.

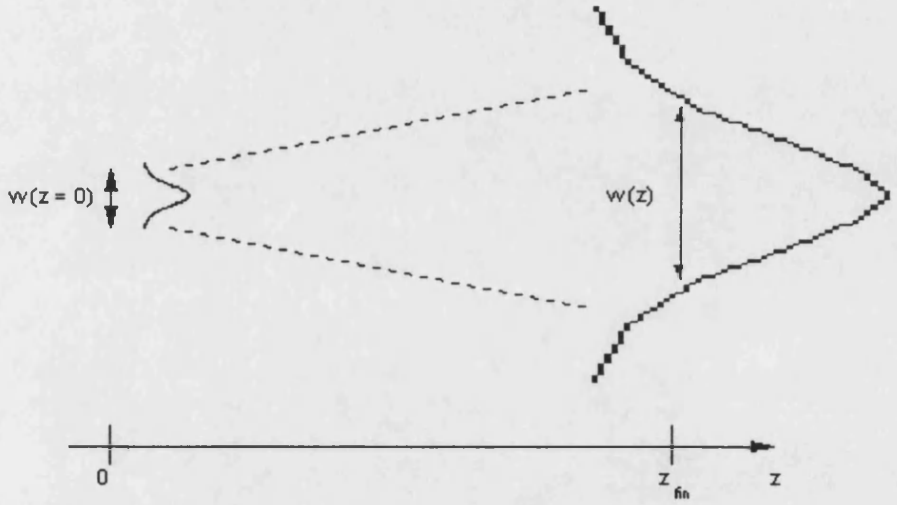


Fig. 5.2.1: Diffraction in half-space: the initial Gaussian field is specified at z = 0, z_{fin} is the propagation length, w(z = 0) = initial beam width, w(z) = beam width at the longitudinal position z.

Consider in the following example an initial Gaussian field distribution propagating in free-space, Fig. 5.2.1. The HGCM is applied to the paraxial wave equation (1.7.6) to solve for the total field diffracting in free-space. The analytic solution is readily available, [Chapter 2, Section 1]. For a further comparison the analytic method of Plane Wave Decomposition [Appendix 1.2] is also used. The results are shown in Fig. 5.2.2. The field profiles obtained with the HGCM and with the Plane Wave Decomposition are in excellent agreement with the analytic solution.

The accuracy in solving free-space (paraxial) propagation thus demonstrates that the Hermite-Gauss total field expansion method is an effective and simplified method for the analysis of diffraction problems, that can be readily applied also to arbitrary initial field distributions. Hence, the HGCM can be considered as very convenient and versatile alternative method to the analytic HGB solutions.

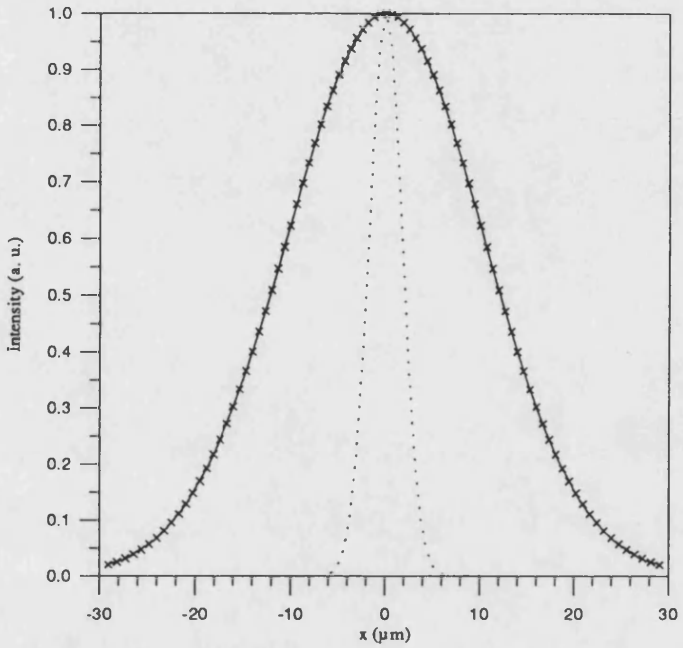


Fig. 5.2.2: Diffraction in half-space (λ_o = 1.55μm, the beam waist of the initial field distribution is w(z = 0) = 5.0μm, the refractive index of the homogeneous medium is n_h = 3.42, z_{fin} = 500μm): dotted line: Gaussian initial field distribution; solid line: HGCM (M = 99, w_o = 2.8μm); x: fundamental Gaussian Beam at z = z_{fin}; broken line: Plane Wave Decomposition method.

2.2 Multimoded symmetric slab dielectric waveguide

More tests to assess the accuracy of the HGCM in solving (total) field propagation are discussed in this section. The problem analysed here is that of calculating the total field propagating in a multimoded lossless dielectric waveguide, given an initial field $F(x, z = 0)$ composed of two bound modes of equal amplitude

$$F(x, z = 0) = a_o f_o(x) + a_2 f_2(x) \qquad (5.2.1)$$

where $f_o(x)$ and $f_2(x)$ are, respectively, the (normalised) fundamental and second order modes supported by the waveguide; and $a_o = a_2$ the corresponding expansion coefficients, which have been chosen to be equal.

Although the contribution of the individual modes does not change during propagation, the field intensity profile can change very significantly because of the relative change in phase due to the different propagation constants characterising the two individual modes. At any longitudinal position the field is

$$F(x,z) = a_o f_o(x)\exp(-i\beta_o z) + a_2 f_2(x)\exp(-i\beta_2 z) \qquad (5.2.2)$$

where β_o and β_2 are the modal propagation constants. The intensity profile (at $z = z_{fin}$) calculated with the HGCM is thus compared with the analytic solution.

This test has been performed in two waveguides characterised by different index steps. The excellent agreement between the HGCM solution and the analytic is demonstrated by Fig. 5.2.3 for the 'stronger' (larger index step) waveguide. In the calculations presented in this section the supergaussian representation of the refractive index has been used. However note that since the waveguide is longitudinally uniform a piecewise constant refractive index distribution would also give accurate results once a suitable w_o is chosen.

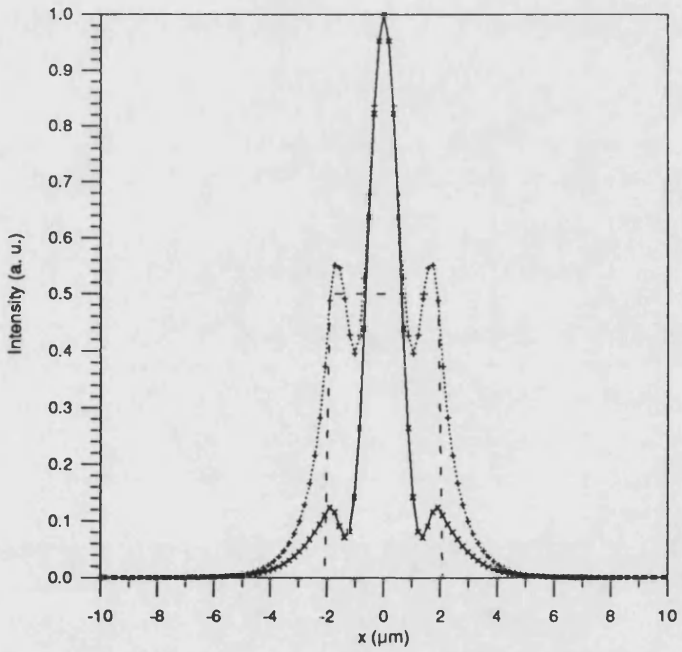


Fig. 5.2.3: Propagation in a multimoded waveguide (λ_o = 0.86 μm, n_1 = 3.33, n_2 = 3.32, $2a_o$ = 4μm, $n_{eff(o)}$ = 3.3289, $n_{eff(1)}$ = 3.3258, $n_{eff(2)}$ = 3.3213, respectively for each bound mode). The results are labelled as follows: dotted line: input field, solid line: HGCM (M = 99, w_o = 0.776μm, z_{fin} = 100μm), (x): analytic solution. The rib is shown as a rectangle (broken line).

The total field calculated with the HGCM has been decomposed in terms of the HG basis set. The magnitude of the expansion coefficients is plotted in Fig. 5.2.4 as a function of the order of the HG functions. The following observations can be made in this respect: i) a smaller number of expansion terms could be used in the calculations since the contribution of the higher order expansion terms is small; however, the large number of expansion terms used in the calculations (M = 99) is justified by the requirement of a detailed field profile, which can be provided only by having a large number of collocation points, and ii) only the symmetric HG functions contribute to the field since the initial field is composed only of symmetric modes.

The close agreement obtained between the HGCM and the analytic solution proves that the HGCM can be effectively used for solving total field propagation problems in longitudinally uniform dielectric waveguides. However, as already noted previously, if the w_o parameter is too large (poor resolution) or too small (the computation window not sufficiently wide) the accuracy decreases.

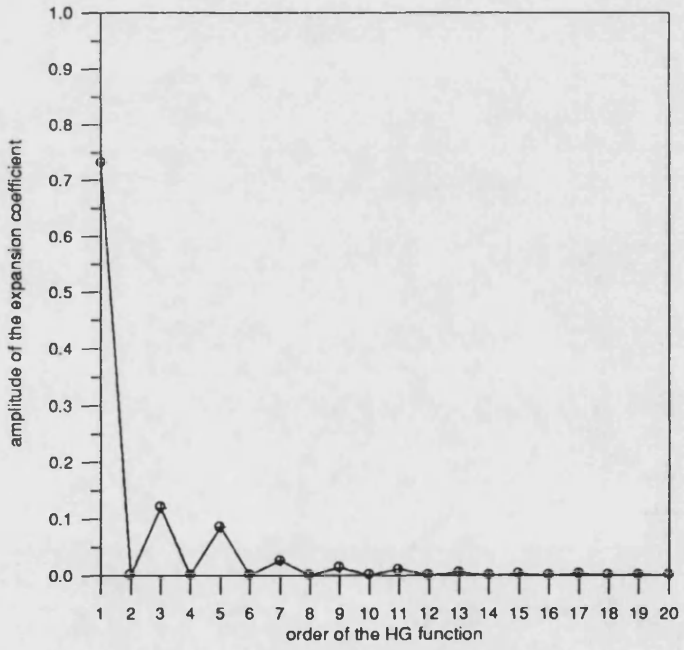


Fig. 5.2.4: Propagation in a multimoded waveguide: contribution of the HG expansion functions to the field calculated in Fig. 5.2.3. (o): coefficients of the HGCM solution, (+): coefficients of the analytic solution. Note that although M = 99 only the coefficients of the HG functions of order < 20 are shown here since the contributions of the HG functions of order > 20 are negligible.

2.3 Coupled waveguides

Propagation in coupled waveguides is also analysed as yet another test of the applicability and the effectiveness of the HGCM. This problem is often solved with coupled mode theory [7]. Here, however, the results calculated with the HGCM are compared with those from the essentially analytic solution derived for modal analysis in a five-layer slab waveguide.

Consider a symmetric structure composed of two identical symmetric waveguides separated by a distance d, Fig. 5.2.5, but which should be rigorously viewed as a five-layer slab waveguide. Fig. 5.2.6 shows the excellent agreement between the field calculated with the HGCM and the analytic solutions. The results are presented for a propagation length corresponding to half the coupling length to emphatically demonstrate mode coupling in the waveguide and thereby to illustrate further the capability of the HGCM.

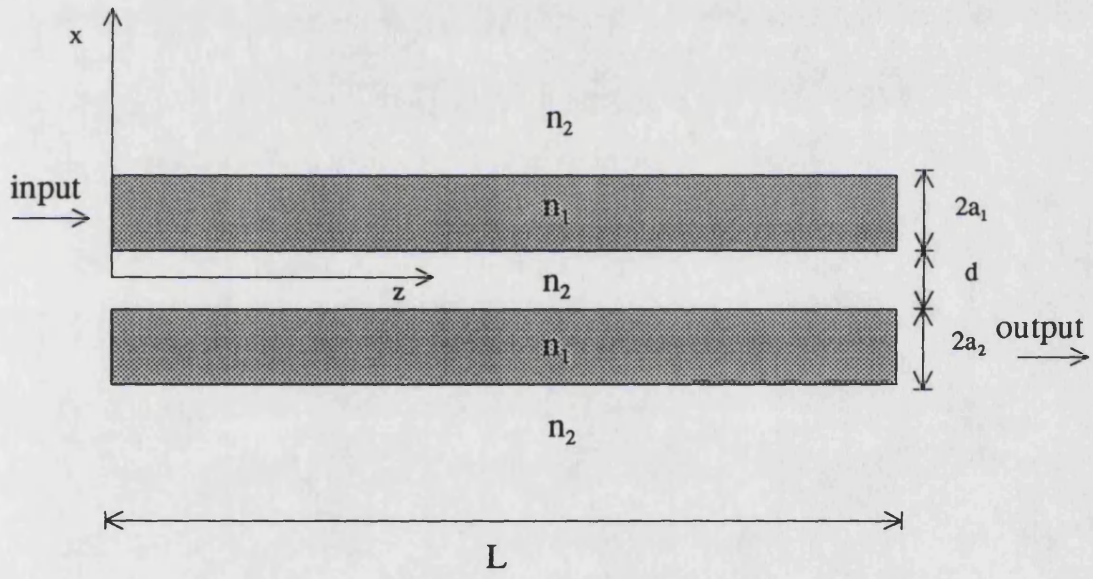


Fig. 5.2.5: Propagation in coupled waveguides (λ_o = 1.55 μm, n_1 = 3.28448, n_2 = 3.28241, $2a_1 = 2a_2$ = 3μm, d = 3μm). The input field is launched in one of the waveguides. The coupling length is $L = 2\pi/\Delta\beta \approx$ 2240μm since the propagation constants[1] of the fundamental and first order modes of the structure are $\beta = k_o n_{eff}$, with $n_{eff(o)}$ = 3.28334 and $n_{eff(1)}$ = 3.28274.

[1] The effective refractive indices of the two modes supported by the five-layer slab waveguide calculated with the HGCM are $n_{eff(o)}$ = 3.28337 and $n_{eff(1)}$ = 3.28271, using M = 99 and w_o = 2.0μm.

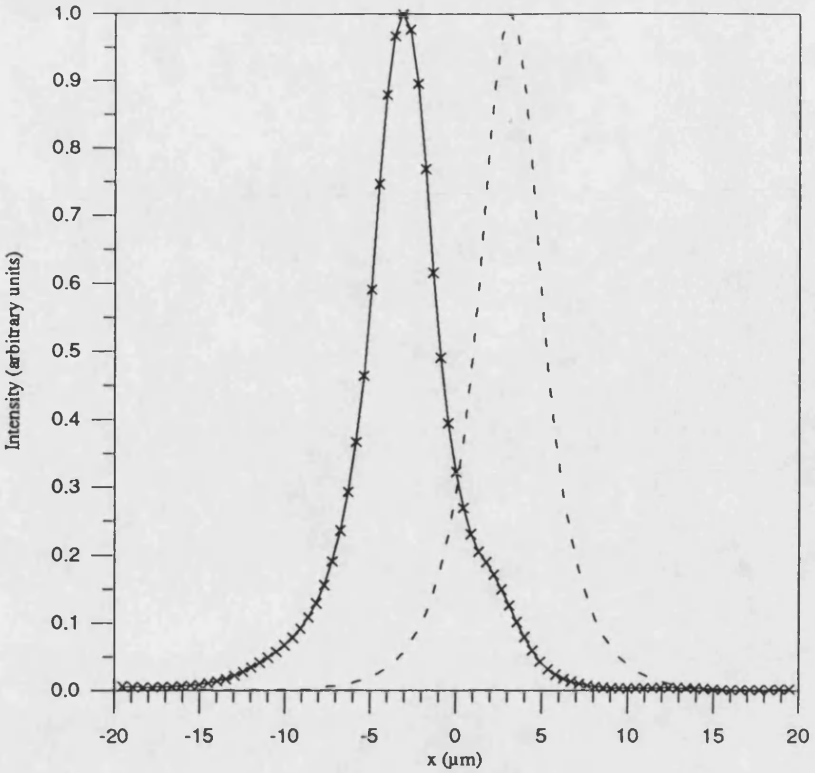


Fig. 5.2.6: Field propagation in coupled waveguides ($z_{fin} = \frac{L}{2}$ = 1120μm, other parameters as shown in Fig. 5.2.5): solid line: HGCM (M = 99, w_o = 2.01μm); crosses: analytic solution; broken line: initial field distribution.

2.4 Longitudinal step discontinuities in waveguides

In this section it is demonstrated that the HGCM is particularly suited to the solution of electromagnetic field propagation in slab dielectric waveguides with longitudinal non-uniformities. A stark example of such non-uniformities is the step discontinuity, Fig. 5.2.7. Referring to the process of field expansion in terms of complete (orthogonal) sets [Chapter 3, Sections 1.2, 2.1 and 2.3] the longitudinal non-uniformity may be formulated as the excitation of the mode-spectra of the uniform waveguide due to the perturbation produced by the discontinuity. It is, of course, not essential to use the waveguide mode spectrum as the basis set; the fundamental point is that the expansion set be complete in the functional space of interest, [Chapter 3, Section 1.2]. Therefore the HG set is used here since it is argued that it provides for a very suitable and convenient set for field analysis in dielectric structures which also accounts for the radiation mode spectrum of the waveguide.

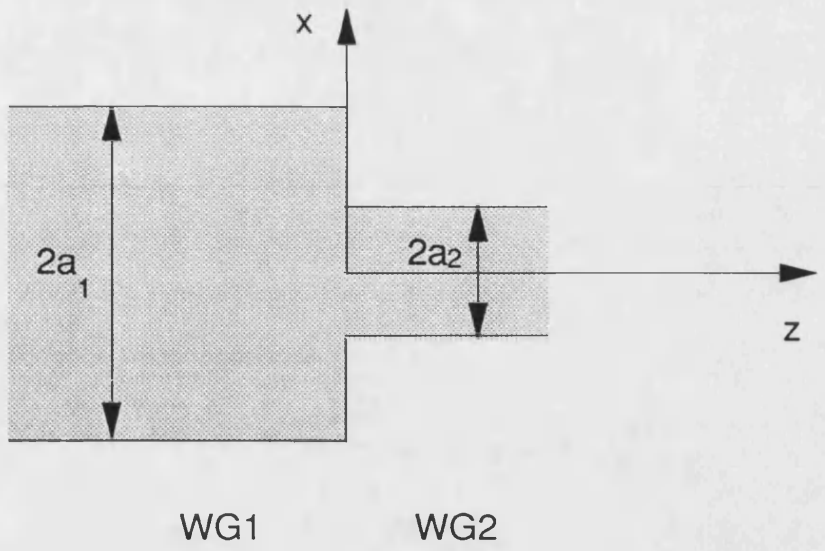


Fig. 5.2.7: Junction between two dielectric waveguides, WG1 and WG2, of width $2a_1$ and $2a_2$, respectively.

Field propagation in step discontinuities has been investigated in detail in the literature, with particular attention to the power lost in radiation. In [8] an analytic method, referred to as step-transition method, based on the mode matching technique is discussed. There the amplitudes of the reflected and transmitted, guided and radiation, modes are determined by matching the transverse components of the electromagnetic field at the junction (specifically, waveguides supporting only one bound mode are considered on both sides of the discontinuity). In [9] and in [10] a mode matching method is proposed in which the dielectric waveguide junction is enclosed in a metal (closed) waveguide in order to avoid the complication of matching radiation mode components. The latter approach is justified in the paper [10] by the one-to-one correspondence of the mode spectra of open and closed guiding structures[1], and by the assumption that in the limit of metal boundaries at infinity the two structures become identical.

In the following example the case of a dielectric structure of the type shown in Fig. 5.2.7 is analysed. The waveguide on the left hand side (WG1, $z < 0$) supports three (bound) modes, while the one on the right hand side (WG2, $z > 0$) supports only

[1] In a closed waveguide the spectrum is composed of an infinite discrete set of eigenvalues, while in an open waveguide the spectrum is composed of a finite discrete part complemented by an infinite continuous component. The surface, radiation and reactive modes of open waveguides correspond, respectively, to the slow, fast and evanescent modes of closed waveguides, [10].

the fundamental mode. Assume propagation along the positive z-axis. At the discontinuity, because of the difference in width the input field launched from WG1 into WG2 couples only partially into the fundamental mode of WG2 and the rest of the field diffracts outside WG2 into the cladding layer. The reflected field is assumed to be negligible, which is justified by the weak refractive index step considered in the present analysis. Consider the three situations:

1) the initial field launched into WG2 is the fundamental mode of WG1. The mismatch due to the widths of the two waveguides produces 'wings' diffracting in the cladding layers (radiation), that gradually depart from the region nearer to the core of WG2 as the field propagates along z, Fig. 5.2.8 (a);

2) the input field is the first order mode of WG1. Since WG2 does not support anti-symmetric modes, the field couples only to the radiation modes and hence the diffracting 'wings' (radiation) are more pronounced than in the previous case, Fig. 5.2.8 (b);

3) the initial field is the second order mode of WG1. This situation is similar to the one described in 1). The field partially couples in the fundamental mode supported by WG2, the rest of the field diffracts into the cladding layers (radiation), Fig. 5.2.8 c).

2.4.1 Approximate analysis of waveguide step-discontinuities

The results calculated with the HGCM are compared with those obtained with an approximate but new method of analysis (referred to as Plane Wave Approximation Spectrum, PWAS) based on the following considerations. In the situations 1) and 3) described above the incident field partially couples into WG2, this portion of the field, $F_1(x,z)$, is then guided in WG2. It is thus possible to consider $F_1(x,z)$ as a separated field from the diffracting 'wings', $F_2(x,z)$. The latter, $F_2(x,z)$, is then solved as a diffracting field in a homogeneous medium chosen to be that of the cladding, using the plane wave expansion method [Appendix 1.2]. Note that $F_2(x,z)$ is only approximately determined since, strictly, it is affected by the presence of the waveguide WG2. However, it is assumed that at any longitudinal position the sum of the two fields $F_1(x,z) + F_2(x,z)$ corresponds to the total field propagating in the region $z > 0$.

Specifically, the initial field distribution, $F_o(x)$, is given. The overlap integral, K, of $F_o(x)$ with the fundamental mode, $f_o(x)$, of WG2 is determined by

$$K = \int_{-\infty}^{+\infty} F_o(x) f_o(x) dx \tag{5.2.3}$$

Hence $F_1(x,z)$ is

$$F_1(x,z) = K f_o(x) e^{-i\beta_o z} \tag{5.2.4}$$

where β_o is the propagation constant of $f_o(x)$.

The field $F_2(x,z)$ at the junction, $z = 0$, is given by

$$F_2(x, z=0) = F_o(x) - F_1(x, z=0) \tag{5.2.5}$$

and at any $z > 0$ $F_2(x,z)$ can be determined using the following plane wave expansion

$$F_2(x,z) = \int_{-\infty}^{+\infty} A(k_x) e^{-ik_x x} e^{-ik_z z} dk_x \tag{5.2.6}$$

with $k_x^2 + k_z^2 = k_o^2 n_2^2$ (n_2 is the refractive index of the cladding layer), and the amplitude distribution given by

$$A(k_x) = \int_{-\infty}^{+\infty} F_2(x, z=0) e^{ik_x x} dx \tag{5.2.7}$$

The agreement between the HGCM and the PWAS method is not expected to be perfect because of the approximations made for the latter method. However, it is satisfying to observe that both methods show good agreement and, particularly, reveal the radiation outside WG2, Fig. 5.2.8.

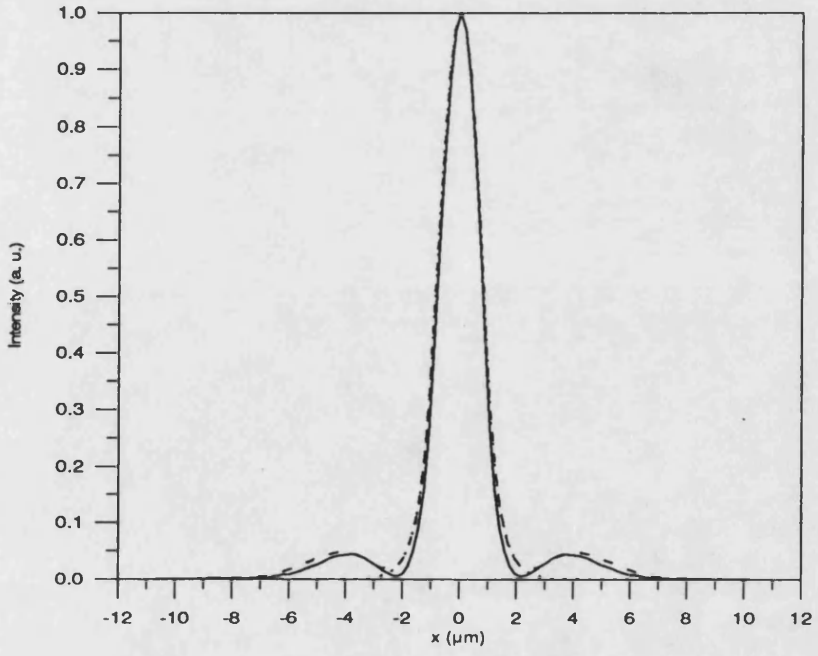
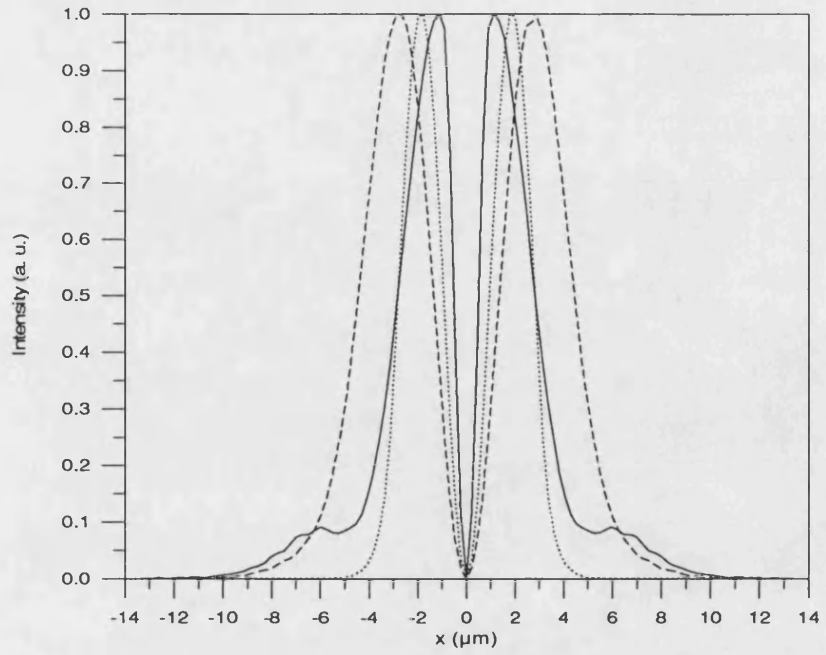
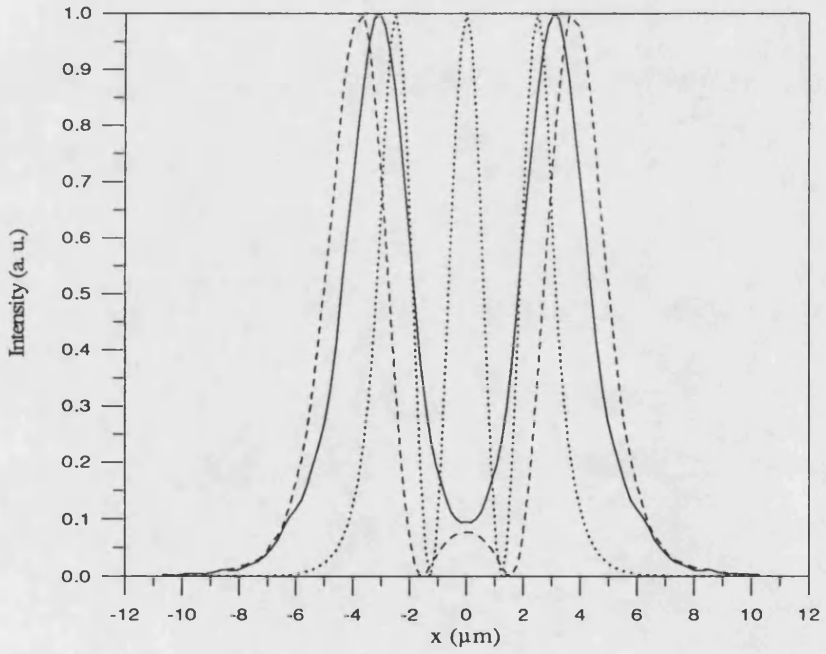


Fig. 5.2.8 (continued): a)



b)



c)

Fig. 5.2.8 (continued): Field propagation across a step discontinuity calculated at distance of 50μm from the junction: dotted line: input field, solid line: HGCM (M = 99, w_o = 1.2 μm), broken line: PWAS method. Parameters: $2a_1$ = 6μm, $2a_2$ = 2μm, n_1 = 3.32, n_2 = 3.3, λ_o=1.55μm. The mode effective refractive indices are, in WG1: $n_{eff(o)}$ = 3.31842, $n_{eff(1)}$=3.31378, $n_{eff(2)}$ = 3.30652, in WG2: $n_{eff(o)}$ = 3.31194.

2.5 Field propagation in tapered dielectric structures

It is particularly for the analysis of field propagation in longitudinally non-uniform dielectric structures that the HGCM has been developed in this research programme. The results obtained by using the HGCM to solve for electromagnetic fields in linear dielectric tapers are presented in this section. A piecewise constant refractive index distribution defines the core/cladding regions and the taper is represented by a core width that is a function of the longitudinal position, i.e., a = a(z), Fig. 5.2.9.

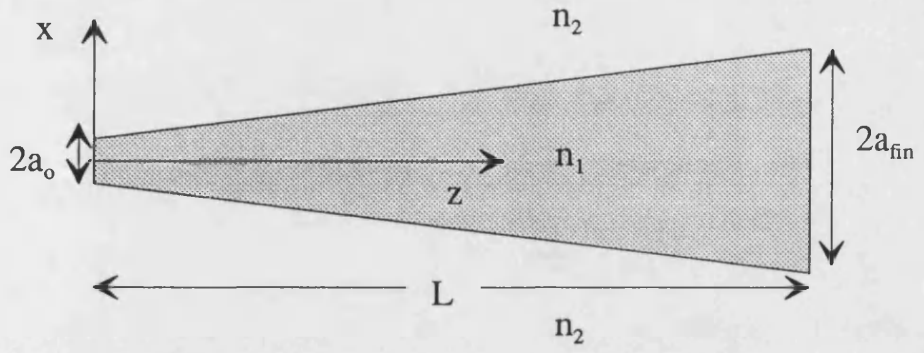


Fig. 5.2.9: Top view of a linearly tapered structure.

The advantages of using a continuous refractive index distribution with the HGCM have been discussed in the previous Section 1. Hence in the present section the linear taper is represented using a sharp, but continuous, supergaussian refractive index distribution similar to the one presented in Section 1. However, to account for the tapering width, equation (5.1.4) is written as

$$n(x,z) = n_2 + (n_1 - n_2)e^{\frac{-1}{2}\left(\frac{x}{a(z)}\right)^q} \qquad (5.2.8)$$

where $a(z) = a_o + t \cdot z$ is the half the width of the taper, q specifies the refractive index profile, and n_1 and n_2 the core and cladding refractive indices far from the transition.

The results obtained with the HGCM are then compared with those from the method described in [1]. The latter method is based on the mode matching technique, except that it uses only the discrete part of the (local) mode spectrum (referred to as DMM), [1]. The results from both methods are shown in Fig. 5.2.10. The agreement is good, but the field calculated with the HGCM is slightly wider than that calculated with the DMM.

It is also noted that with the HGCM the integral of the electromagnetic (optical) intensity distribution, which is a measure of the power carried by the field, is conserved. In fact, using the HGCM it is found that

$$\int_{x_1}^{x_M} |F(x, z = 0)|^2 \, dx \approx \int_{x_1}^{x_M} |F(x, z = z_{fin})|^2 \, dx \qquad (5.2.9)$$

to within an error of 0.01%, where (x_1, x_M) is the computational range set by the collocation points, and F(x,z) is the computed field at any $z > 0$. On the other hand, with the DMM it is found that the error in (5.2.9) is of the order of 10%. It is argued that this discrepancy represents an indication of the contribution of the (local) radiation mode spectrum, which is neglected in the DMM, but implicitly included in the HGCM.

Two other results are shown in Fig. 5.2.10. One is the plot of the local fundamental mode at the output facet. The difference between the latter field and the HGCM computed field propagating from the narrow to the broad output facet shows that the original input fundamental mode couples significantly into higher order local modes even for a quite slowly varying taper.

The second plot reproduces the field profile to be expected at $z = L$ if the same initial field is diffracting into the homogeneous half-space defined by n_2. This result establishes that the weak but finite dielectric step, $\Delta n = n_1 - n_2$, of the taper does affect the field shape.

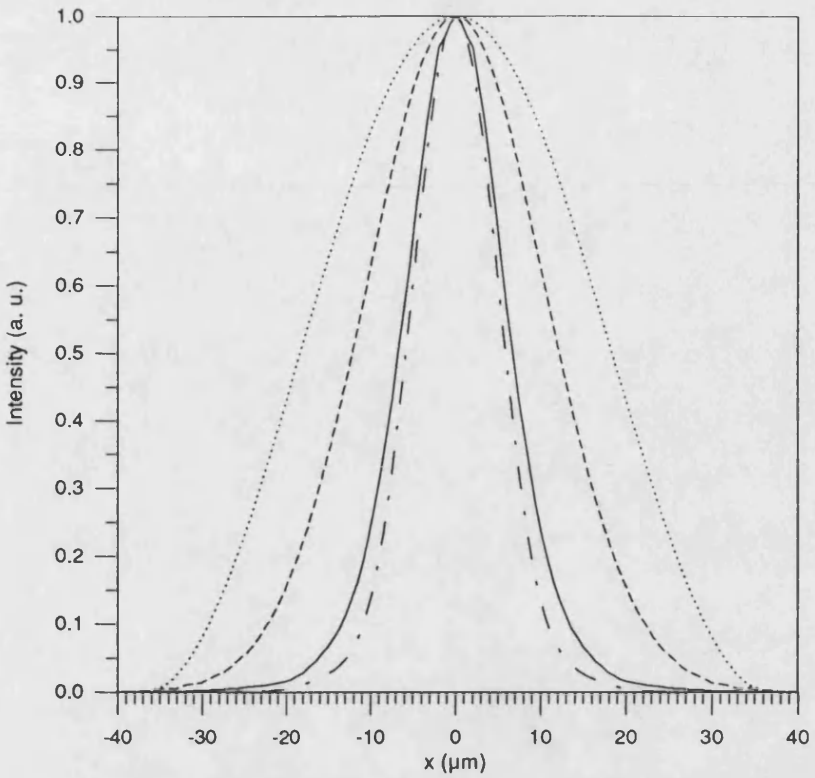


Fig. 5.2.10: Optical field propagating in a tapered structure, the input field is the (local) fundamental mode at z = 0 (n_1 = 3.42, n_2 = 3.4187, $2a_o$ = 5μm, $2a_{fin}$ = 70μm, L = 500μm, λ_o = 1.55μm): solid line: output field calculated with the HGCM (M = 99, five longitudinal sections were used with w_o = 0.9 - 1.2 - 1.8 - 2.6 - 3.0μm, respectively); broken-dotted line: output field calculated with the DMM; dotted line: (local) fundamental mode at the broad end of the taper; broken line: free-space diffraction of the same initial field (as in Fig. 5.2.1).

2.6 Parabolic structures

There is considerable interest in optoelectronics to improve the device characteristics in order to maximise the coupling between components that have different sizes or shapes, e.g., coupling semiconductor lasers to slab waveguides or to optical fibres. Tapered structures are thus very suited to this purpose; however, to minimise the radiation losses and the coupling to higher order (local) modes, some requirements need to be satisfied. In this respect various tapering profiles have been investigated as discussed below.

In [11] the optimal tapering shape is sought to minimise radiation power losses along the length of the device. It is proved therein that adiabatic operation for mode transformers can be achieved when the lowest order local normal mode propagates through the structure with minimum coupling to the higher order local modes. An

'adiabaticity' criterion is thus given, based on the consideration that the 'spreading of the waveguide width' should be slower than the diffraction spreading of the lowest order mode, [11].

Theoretical modelling shows that mode conversion to higher order modes increases in structures with rapidly changing taper widths, and hence longer (slowly varying) tapers may be used to reduce this problem. However, practical difficulties may be encountered in the realisation of long tapers; shorter tapers are often desired for many applications, for example in integrated optics.

It has been demonstrated, [11], that parabolic tapers satisfy the adiabaticity criterion, such that it is possible to retain almost the total power in the fundamental local mode all along the length of the taper. Parabolic structures of the type shown in Fig. 5.2.11 may thus be better suited for applications where mode size increase is required but without conversion to higher order modes.

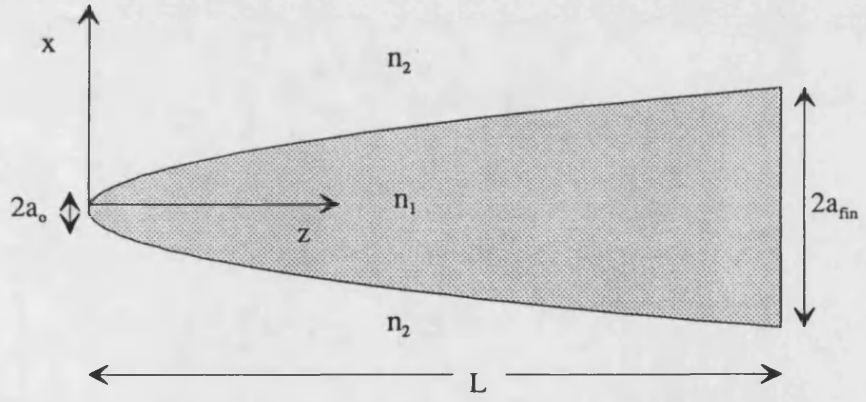


Fig. 5.2.11: Parabolic taper, the width of the taper varies along z as $a^2(z) = a^2_o + pz$, where p is the tapering parameter ($2a_o$ is the input width, $2a_{fin}$ is the output width, L is the length of the taper).

The HGCM is applied to solve for field propagation in parabolic tapers. Two types of comparison are possible. First the results calculated with the HGCM may be compared with those calculated by the DMM, [2]. Second, the results calculated with the HGCM for parabolic tapers can be compared with those obtained for corresponding linear tapers, i.e., of the same dimensions.

Two sets of results are shown in Fig. 5.2.12. The results obtained for the parabolic taper are shown in Fig. 5.2.12 a), while those obtained for the

corresponding linear taper are plotted in Fig. 5.2.12 b). In particular, in each set three curves are shown, of which two have been computed with the HGCM (one using a piecewise constant, the other a continuous supergaussian refractive index profile), while the third has been calculated with the DMM.

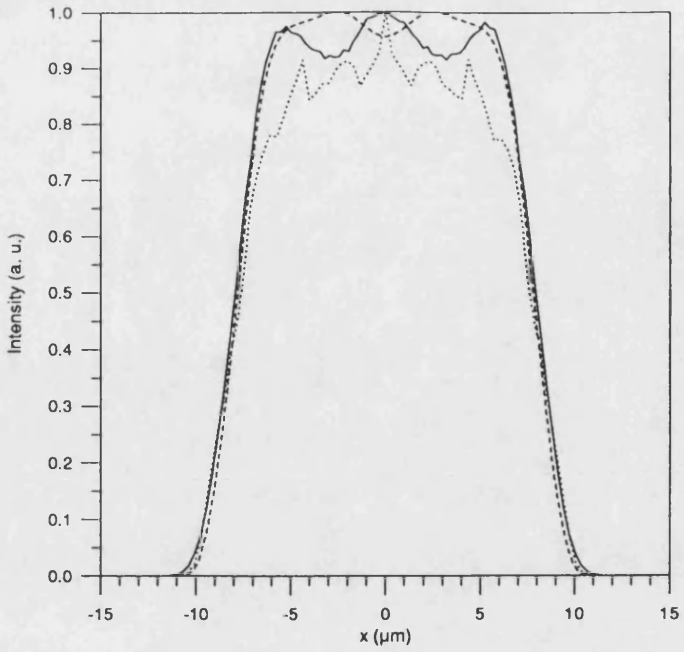
From the comparison of Fig. 5.2.12 a) and b) it can be observed that

i) as expected, the use of a SG refractive index profile greatly improves the quality of the results calculated with the HGCM;

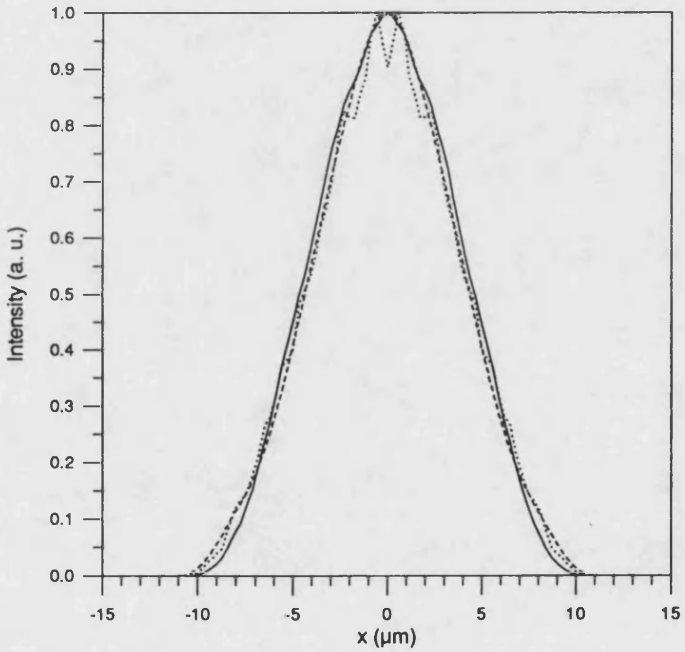
ii) the near field calculated for the linear and the parabolic structures are significantly different;

iii) each set of results calculated with the HGCM is in good agreement with the corresponding set calculated with the DMM.

The corresponding phase fronts (at the output facet) are also shown in Fig. 5.2.13 a) and b), form which it emerges that a characteristic feature of parabolic tapers is to produce a flatter phase front (compared to the linear taper), which yields a narrower far field, Fig. 5.2.14 a) and b).



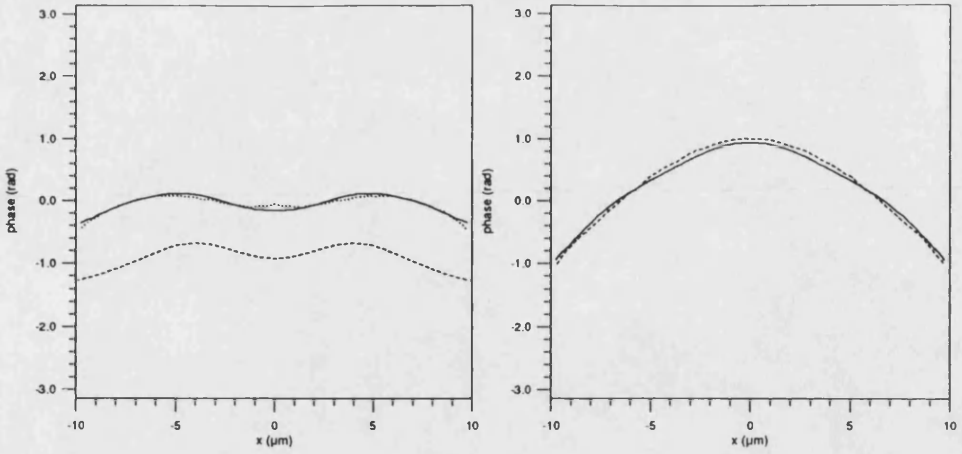
a)



b)

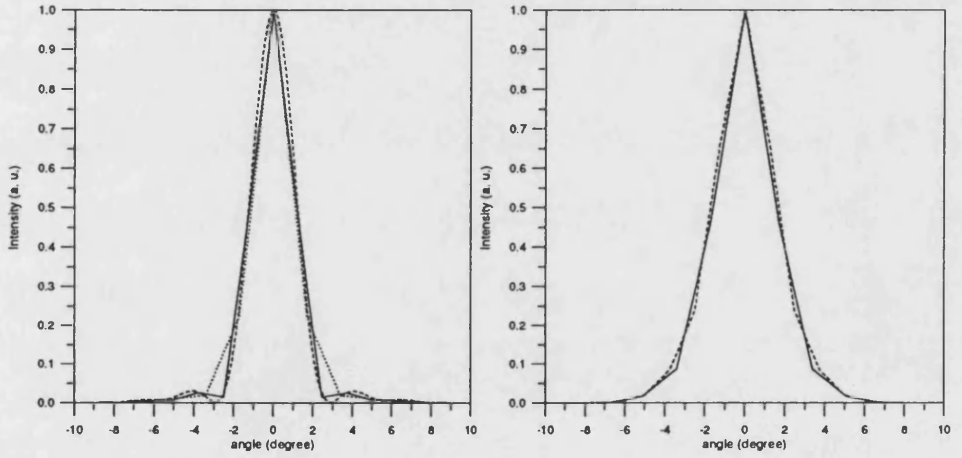
Fig. 5.2.12: Field propagation in a) parabolic and b) linear tapers ($2a_o$ = 3μm, $2a_{fin}$ = 20μm, L = 500μm, n_1 = 3.33, n_2 = 3.32, λ = 0.86μm), solid line: SG (q = 16); dotted line: ST broken line: DMM.

Parameters used for the HGCM (M = 99, Δz = 5μm): parabolic taper: 5 longitudinal sections w_o = 0.3 - 0.5 - 0.7 - 0.9 - 1.1 μm, with $z_{section}$ = 40 - 80 - 120 - 200 - 500 μm; linear taper: 3 longitudinal sections w_o = 0.9 - 1.1 - 1.3 μm, with $z_{section}$ = 100 - 300 - 500 μm.



a) b)

Fig. 5.2.13: Phase front at the output facet of a) parabolic, b) linear taper: solid line: SG (q = 16); dotted line: ST Linear taper: broken line: DMM. (Parameters in Fig. 5.2.12.)



a) b)

Fig. 5.2.14: Far field from the output facet of a) parabolic, b) linear taper: solid line: SG (q = 16); dotted line: ST Linear taper: broken line: DMM. (Parameters in Fig. 5.2.12.)

2.7 Sensitivity to w_o and M

As with the solution of the eigenvalue problems, the solution of initial value problems by the HGCM are also improved by the use of a sharp, but continuous dielectric distribution [Section 1 in this chapter]. Hence it is expected that the choice of the width parameter of the HG expansion functions, w_o, is decided primarily by the extent of the structure to be modelled.

The effect of the number of expansion terms, M (which in the HGCM corresponds to the number of collocation points) on the results calculated with the HGCM is

studied in this section in the context of the initial value problem. Hence field propagation in a linear taper is solved here with the HGCM considering M as a variable parameter. Both the piecewise constant (ST) and the supergaussian (SG) refractive index distributions have been used in the calculations. The results are shown in Fig. 5.2.15 and Fig. 5.2.16 for the ST and the SG dielectric profile, respectively. Different longitudinal sections, each characterised by a particular value of w_o, have been used in the simulations. The w_o values have been chosen so that with varying M the same computational window is used in all cases.

From the comparison of Fig. 5.2.15 and Fig. 5.2.16 it is evident that convergence is not easily achieved when a piecewise constant refractive index profile is used in the HGCM calculations. On the other hand, by using a continuous refractive index profile the accuracy of the HGCM increases even when using a small number of collocation points.

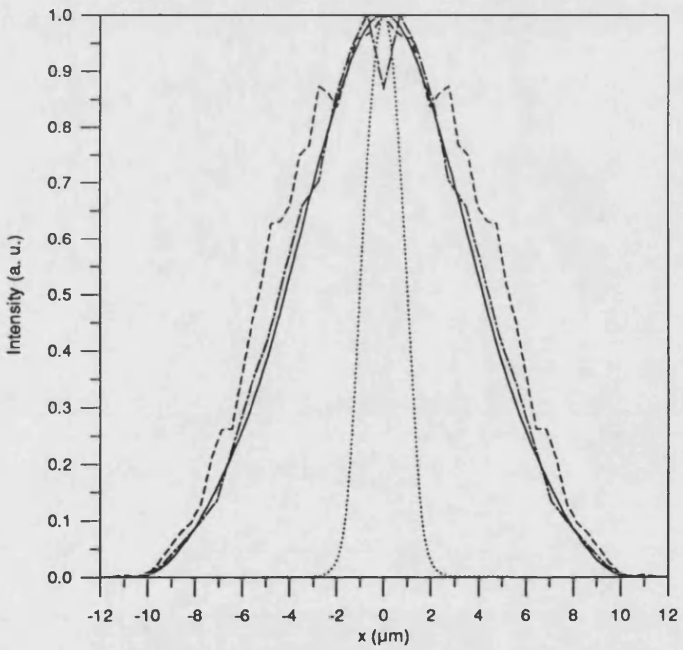


Fig. 5.2.15: Convergence properties of the HGCM with respect to the parameter M, using a ST refractive index distribution: dotted line input field, broken-dotted line: M = 41, dotted line: M = 71, solid line: M = 99. Parameters for HGCM: M = 99, 3 longitudinal sections with $z_{section}$ = 100 - 300 - 500 μm, Δz = 5μm (for M = 41: w_o = 0.7 - 1.2 - 1.8μm; for M = 71: w_o = 0.7 - 1.2 - 1.5μm; for M = 99: w_o = 0.4 - 0.9 - 1.2μm). Structure as in Fig. 5.2.12.

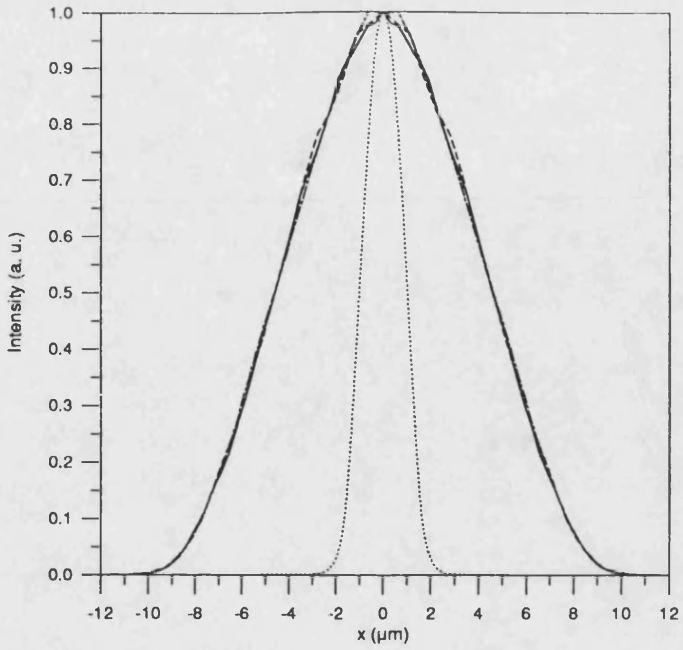


Fig. 5.2.16: Convergence properties of the HGCM with respect to the parameter M, using a SG refractive index distribution: dotted line input field, broken-dotted line M = 41, dotted line M = 71, solid line M = 99. (Parameters in Fig. 5.2.17.)

For the same structure, it has been observed that it is possible to use a longitudinal step as large as $\Delta z = 10\mu m$ without noting changes in the results calculated with the HGCM (typically $\Delta z = 5\mu m$ is used in this thesis); for larger Δz the accuracy decreases.

Tab. 5.2.1 compares the computation times needed for the two methods, HGCM and DMM, to solve for the field propagation problem discussed above. It is clear that the comparison is not completely justified since the DMM only takes into account bound modes, while the HGCM is a total field analysis. It is noted that the computing time needed for the HGCM increases approximately as M^2.

Tab. 5.2.1: Comparison of the computing time needed for HGCM and DMM to solve initial value problems. In particular the CPU time (minutes: seconds) is given as a function of M for the HGCM.

HGCM		DMM
M	CPU	CPU
41	0:36	0:30
51	1:00	
61	1:30	
71	2:41	
81	4:14	
91	6:13	
99	9:16	

Summary

The HGCM has been demonstrated to be most suited to solve for a variety of field problems in dielectric structures. The following conclusions can be drawn from the analysis of the results presented in this chapter:

i) the HGCM can be used to solve both eigenvalue and initial value problems for fields in open structures;

ii) the eigenvalue problem for any real refractive index distribution may be accurately solved by the HGCM. Complex refractive index profiles may also be solved with the HGCM with satisfactory accuracy (for small imaginary parts of the refractive index in the core layer);

iii) the use of a continuously varying refractive index distribution in the calculations dramatically increases the accuracy of the HGCM. It has been observed that by using analytic representations of the refractive index profile the choice of w_o may be left almost arbitrary;

iv) the HGCM has been used for solving field propagation in the form of an initial value problem. In most of the cases analysed the results produced with the HGCM are in good agreement with those obtained from other methods of solution. The HGCM has been found to be particularly useful in the solution of problems involving radiation (e.g., diffraction in homogeneous media, propagation across step discontinuities). The consideration of the total field expressed on the basis of the complete HG set is

regarded as an advantage if the HGCM is to be compared with a solution method based on the commonly used mode matching method since it is expected that the inclusion of the radiation modes in the field matching complicates the formalism.

Appendix 5.1

Variable transformation in the collocation method

In some cases it is convenient to redistribute the collocation points along the transverse domain in order to increase the density of sampling points. This can be achieved by means of a transformation of variable in the collocation method, [6].

Consider a variable transformation of the form x = g(s). Two conditions must be satisfied in order to apply it to the collocation method: i) g continuous, with first and second derivatives also continuous on the transverse range, and ii) $g'(s) > 0$. The field in the dielectric structure is thus written as

$$F(x,z) = \sqrt{g'(s)}\,E(s,z) \qquad \text{(A5.1 - 1)}$$

Substituting (A5.1 - 1) into the scalar wave equation (1.4.2), obtain the differential equation for E(s,z)

$$\partial_z^2 E(s,z) + C_1(s)\partial_s^2 E(s,z) + \left[C_2(s) + k_o^2 \varepsilon(s,z)\right]E(s,z) = 0 \qquad \text{(A5.1 - 2)}$$

with $C_1(s) = (g'(s))^{-2}$ and $C_2(s) = \dfrac{2g'''(s)g'(s) - 3(g''(s))^2}{4(g'(s))^4}$.

Equation (A5.1 - 2) is now solved using the collocation method with respect to the variable s. This means that it is now E(s,z) to be expanded in terms of the HG basis set

$$E(s,z) = \sum_{k=1}^{N} B_k(z) f_k(s) \qquad \text{(A5.1 -3)}$$

with $B_k(z)$ the z-dependent expansion coefficients and $f_k(s)$ the HG functions, equation (2.2.4). The total field at the redistributed collocation points, $\{x_i\}_{i=1,2,\ldots,N}$, is thus obtained using equation (A5.1 - 1).

References

[1] I. Middlemast, J. Sarma, P. S. Spencer, *Characteristics of tapered rib-waveguides for high-power semiconductor optical sources*, IEE Proceedings-Optoelectronics, 144 (1997), pp. 8-13

[2] N. S. Brooks, J. Sarma, I. Middlemast, *A compact model of tapered geometry semiconductor optical devices*, Semiconductor and Integrated Optoelectronic Conference (SIOE '96), paper 45, abstract, Technical Digest, Cardiff, April 1996

[3] A. Sharma, S. Banerjee, *Method for propagation of total fields or beams through optical waveguides*, Optics Letters, 14 (1989), pp. 96-98

[4] A. Sharma, *Collocation method for wave propagation through optical waveguiding structures*, in Progress in Electromagnetics Research, Electromagnetic Waves PIER 11, J. A. Kong ed., EMW Publishing (1995)

[5] M. Born, E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*, Pergamon Press (1986)

[6] A. Sharma, A. Taneja, *Variable transformed collocation method for field propagation through waveguiding structures*, Optics Letters, 17 (1992), pp. 804-806

[7] A. Yariv, *Optical Electronics*, Reinhart and Winston (1985)

[8] D. Marcuse, *Radiation losses of tapered dielectric slab waveguides*, The Bell System Technical Journal, 49 (1970), pp. 273-290

[9] P. H. Masterman, P. J. B. Clarricoats, *Computer field-matching solution of waveguide transverse discontinuities*, IEE Proceedings, 118 (1971), pp. 51-63

[10] G. H. Brooke, M. M. Z. Kharadly, *Step discontinuities on dielectric waveguides*, Electronics Letters, 12 (1976), pp. 473-475

[11] A. F. Milton, W. K. Burns, *Mode coupling in optical waveguide horns*, IEEE Journal of Quantum Electronics, QE-13 (1977), pp. 828-835

Chapter 6

The carrier diffusion equation

One of the main directions of the research work presented in this thesis has been to achieve a compact scheme for the modelling of semiconductor optical devices. The work presented in the chapters so far has concentrated on the electromagnetic (optical) field aspects of the problem by developing the HGCM for the analysis of field propagation particularly in dielectric structures that may be considered as supporting weakly diffracting signals. Indeed the HGCM has been used to solve for media in which gain or loss is introduced formally through a complex dielectric distribution. However, in many practical semiconductor devices the optical gain is produced by creating an population inversion through injected carriers.

Hence, for a complete and consistent model of the device it is not only the field profile but also the carrier distribution that must be analysed. The carrier distribution is obtained by solving the carrier diffusion equation which is typically nonlinear because of the presence of explicit nonlinear effects, such as bimolecular and Auger recombination. In active semiconductor optical devices the situation is further complicated by the possible strong, implicit nonlinearities induced by stimulated recombination. The need for a numerical solution is thus evident.

The contents of this chapter will demonstrate that a novel approach of using essentially the HGCM to solve for the nonlinear diffusion equation provides for a very convenient and efficient technique for computing the carrier distribution.

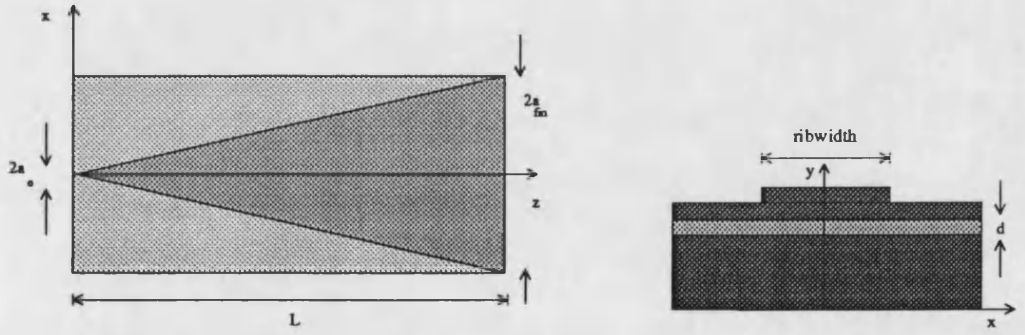
The formalism of the HGCM for the diffusion equation is presented in Section 1. The details of the collocation method used to numerically solve the diffusion equation are illustrated in Section 2. Results from the HGCM are thus presented in Section 3 and compared with those from other numerical techniques. In particular, one set of results refer to the linear form of the diffusion equation which can be solved analytically and hence HGCM numerically computed results for these cases are compared with the analytic solutions. Other results referring to the nonlinear diffusion equation are compared with those from the Jacobi-Tridiagonal method (JTM) described in [1] (in the same reference the JTM is compared with two other numerical methods of solution). Numerical aspects concerning the sensitivity of the HGCM on

the HG width parameter, w_o, and on the number of collocation points, M, in solving the diffusion equation are discussed in Section 4.

1. Formulation

A schematic of a typical tapered geometry semiconductor device is shown in Fig. 6.1.1. In the present context the tapered geometry merely defines the region of current injection, but the geometry is not fundamental to the method of solution. It is assumed that the taper is slowly varying along the longitudinal (z) direction, hence the carrier diffusion along z is neglected so that only diffusion along the transverse (x) axis is analysed in the present formulation.

Note that the tapered contact width varies considerably along the length, so diffusion also has considerably different effects with longitudinal position on the carrier distribution. For example, in the narrower part of the taper the carrier profile tends to be Gaussian-like, while at the wider end it is almost unaffected by diffusion, thereby resulting in a flat-top profile. It is thus required that any solution scheme produce accurate results for the carrier distribution over the full range of widths of a typical tapered geometry device. It is shown in this chapter that the HGCM fulfils this role most satisfactorily.



a) b)

Fig. 6.1.1: Tapered geometry device: a) top view: the shaded region defines the contact for current injection; b) front view: d is the thickness of the active layer.

At any longitudinal position, z, the (rate) carrier diffusion equation, [2], [3] is

$$D\frac{d^2N(x;z)}{dx^2} - B_r\left[n_o + N(x;z)\right]N(x;z) - \gamma N^3(x;z) - g(x;z)P(x;z) + \frac{J(x;z)}{qd} = 0 \quad (6.1.1)$$

where z is considered as a parameter, hence the notation v(x;z) for any variable v(x,z) in the equation, D is the diffusion coefficient, B_r the bimolecular recombination constant, n_o the doping density, γ the Auger recombination coefficient, g(x;z) the local gain profile, P(x;z) the photon density distribution, J(x;z) the injection current profile, q the electronic charge, d the thickness of the active layer (along the y-direction) and N(x;z) the unknown carrier distribution.

Without satisfying self-consistency the stimulated emission term, g(x;z)P(x;z), does not introduce any nonlinearities in the diffusion equation, and hence it is neglected in the present analysis. Hence, the diffusion equation to be solved here is

$$D\frac{d^2N(x)}{dx^2} - B_r[n_o + N(x)]N(x) - \gamma N^3(x) + \frac{J(x)}{qd} = 0 \qquad (6.1.2)$$

where, for brevity, all the z dependent variables v(x;z) have been written as v(x) since z is considered as a parameter in the present context.

It is proposed and demonstrated in this chapter that equation (6.1.2) may be conveniently solved using essentially the same HGCM formalism that has already been applied to solve the optical field equation, [Chapter 4]. Thus, the carrier profile, N(x), is expanded in terms of a complete set of (orthogonal) basis functions, $f_k(x)$

$$N(x) = \sum_{k=0}^{M} a_k f_k(x) \qquad (6.1.3)$$

with $M = \infty$ for the ideal case of a series expansion, and a_k the (constant) expansion coefficients. The boundary conditions typically associated with equation (6.1.2) for structures that are unbounded along x are

$$N(x) \to 0 \quad \text{and} \quad \frac{dN}{dx} \to 0 \quad \text{for} \quad |x| \to \infty \qquad (6.1.4)$$

For the cases of present interest the solution is in the vector space of square integrable functions which are defined by $\int |N(x)|^2 dx < \infty$. The set of functions, $\{f_k(x)\}$, used in the expansion (6.1.3) then must be complete in the same functional space, [4].

Based on features described and discussed in Chapter 4, it has been found most convenient to choose the HG functions as the basis set, $\{f_k(x)\}$, equation (2.2.4), to solve also for equation (6.1.2). This choice seems very appropriate because the HG functions form a complete set in the functional space of interest and also because they individually satisfy the boundary conditions (6.1.4) associated with equation (6.1.2).

Hence the boundary conditions (6.1.4) do not appear explicitly in the subsequent formulation since they are built into the solution as expressed in (6.1.3). A further advantage in using the HG set is that the resulting scheme naturally leads to the use of the numerical collocation method. This aspect has already been discussed in Chapter 4 and will be re-iterated in the next section in the context of the diffusion equation.

The next step is to substitute (6.1.3) into equation (6.1.2). However, the quadratic and cubic nonlinear terms pose difficulties in the proposed method for solving (6.1.2). It is therefore necessary to approximate equation (6.1.2) to an equivalent linearised form and to thereby solve the nonlinear diffusion equation by making use of an iterative procedure. Equation (6.1.2) is thus written in the following form

$$D\frac{d^2N(x)}{dx^2} - B_r\left[n_o + \hat{N}(x)\right]N(x) - \gamma\hat{N}^2(x)N(x) + \frac{J(x)}{qd} = 0 \qquad (6.1.5)$$

where the carrier profile calculated in the previous iteration, $\hat{N}(x)$, is now known.

By substituting (6.1.3) into equation (6.1.5), and using the properties of the HG basis functions, equations (2.2.6) and (2.2.7), a set of coupled equations in the unknown expansion coefficients, a_k, of (6.1.3) are derived, Appendix 6.1. This format, however, need a laborious orthogonalisation process which typically requires extensive numerical integration and is considered to be computationally inefficient. In the present model the numerical integration is replaced by the collocation method, [Chapter 4], which is very efficient. Note, however, that the equivalence between the collocation method and the analytic orthogonalisation procedure can be demonstrated by the Gaussian Quadrature formula for the Hermite polynomials, [5], Appendix 6.1.

2. The HGCM for the solution of the diffusion equation

Assume the solution of equation (6.1.5) to be expressible in the form (6.1.3). In practical cases, the number of expansion terms is finite, i.e., $M < \infty$. Also, for a numerical solution, the independent variable (x) is discretised into a number M[1] of points, x_i with i = 1, 2, ..., M, which are known as the collocation points. The discretisation process implies that the carrier profile, N(x), is sampled at the

[1] It is convenient to choose the number of collocation points to correspond to the number of expansion terms used in (6.1.3), [Chapter 4, Section 4].

collocation points, x_i. The best choice of the collocation points is the set of M zeros of the Hermite polynomial $H_M\left(\frac{x}{w_o}\right)$ when using the HG basis set in (6.1.3).

The expression for the carrier density at the collocation points,

$$N(x_i) = \sum_{k=0}^{M} a_k f_k(x_i) \tag{6.2.1}$$

may be interpreted as the i-th component of an M-dimensional vector $\underline{N}$ which thus represents the carrier transverse profile sampled at the M collocation points. By applying the collocation method, [Appendix 6.2], in conjunction with the HG function expansion technique, referred to as HGCM, equation (6.1.5) is reduced to an algebraic matrix equation in $\underline{N}$ of the type

$$\underline{\underline{T}}\,\underline{N} + \underline{J} = 0 \tag{6.2.2}$$

where $\underline{\underline{T}}$ [defined in Appendix 6.2] is an invertible matrix which contains the carrier profile $\hat{\underline{N}}$ calculated in the previous iteration. The solution of (6.2.2) is used in the next iteration loop, and the iteration process is stopped when convergence is achieved, as schematically illustrated in Fig. 6.2.1.

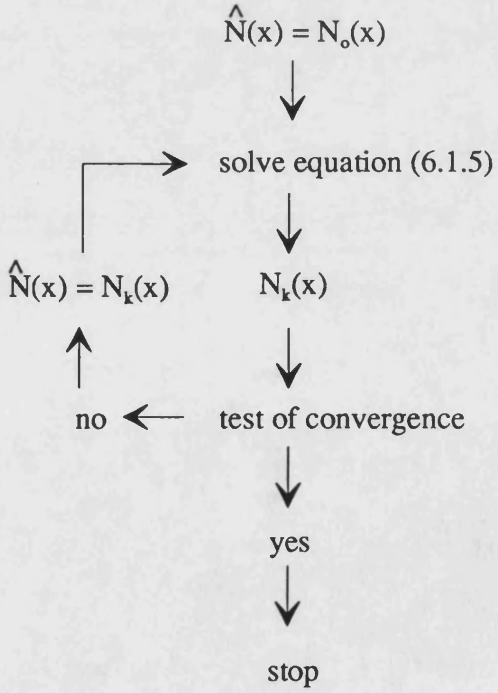


Fig. 6.2.1: Iteration scheme for the solution of the nonlinear diffusion equation.

3. Results

3.1 Linear diffusion equation - comparison with the analytic solution

Three examples of the linear form of the diffusion equation that admit analytic solutions are used to test the HGCM solution scheme. Results of solutions in finite regions, including surface recombination, solved with the HGCM are presented in the following subsections.

The linear diffusion equation considered in the present section is of the form

$$D\frac{d^2N(x)}{dx^2}-\frac{N(x)}{\tau}+\frac{J(x)}{qd}=0 \qquad (6.3.1)$$

where τ is the spontaneous recombination time constant (in s), $J(x)=\begin{cases} J_o & x_1<x<x_2 \\ 0 & x<x_1, x>x_2 \end{cases}$ is the injection current density profile, and the other symbols are as defined in connection with equation (6.1.1).

3.1.1 Unbounded region

The boundary conditions that define the problem are

$$N(x)\to 0 \quad \text{and} \quad \frac{dN}{dx}\to 0 \quad \text{for} \quad |x|\to\infty \qquad (6.3.2)$$

The analytic solution of (6.3.1) takes the form

$$N(x)=\begin{cases} B_1 e^{\frac{x}{\sqrt{D\tau}}} & x<x_1 \\ C_1 e^{-\frac{x}{\sqrt{D\tau}}}+C_2 e^{+\frac{x}{\sqrt{D\tau}}}+\frac{J_o\tau}{qd} & x_1<x<x_2 \\ D_1 e^{-\frac{x}{\sqrt{D\tau}}} & x>x_2 \end{cases} \qquad (6.3.3)$$

where B_1, C_1, C_2, D_1 are constants which are determined by applying the continuity of $N(x)$ and of $\frac{dN(x)}{dx}$ at the points of discontinuity, $x = x_1$ and $x = x_2$, [Appendix 6.3].

The excellent agreement between the carrier profiles calculated with the HGCM and with the analytic solution is demonstrated in Fig. 6.3.1.

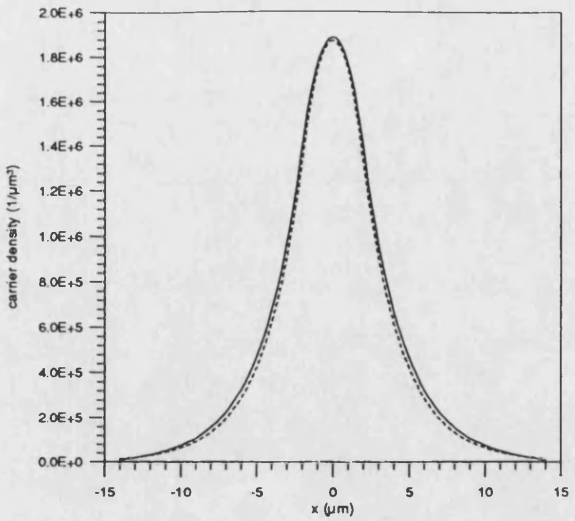


Fig. 6.3.1: Linear diffusion equation in unbounded region (contact width = 4μm, D = 3.5 10^9 μm²/s, τ = 8ns, d = 0.1μm, J_o = 3.0 10^{-5} A/μm²): solid line: HGCM (M = 99, w_o = 1.05μm), broken line: analytic solution.

3.1.2 Finite region

With reference to Fig. 6.3.3, the boundary conditions that have been chosen to be associated with the solution of the diffusion equation in a finite region are

$$N(x_q) = N_q \quad \text{for} \quad x_q = x_o, x_3 \tag{6.3.4}$$

The analytic solution of (6.3.1) is

$$N(x) = \begin{cases} B_1 e^{\frac{-x}{\sqrt{D\tau}}} + B_2 e^{\frac{x}{\sqrt{D\tau}}} & x_o < x < x_1 \\ C_1 e^{\frac{-x}{\sqrt{D\tau}}} + C_2 e^{\frac{+x}{\sqrt{D\tau}}} + \frac{J_o \tau}{qd} & x_1 < x < x_2 \\ D_1 e^{\frac{-x}{\sqrt{D\tau}}} + D_2 e^{\frac{x}{\sqrt{D\tau}}} & x_2 < x < x_3 \end{cases} \tag{6.3.5}$$

where B_1, B_2, C_1, C_2, D_1, D_2 are constants determined by applying the specified boundary conditions (6.3.4) and continuity of N(x) and of $\frac{dN(x)}{dx}$ at $x = x_2, x_3$, [Appendix 6.3]. The solution obtained with the HGCM reproduces the analytic solution with sufficient accuracy, as shown in Fig. 6.3.2.

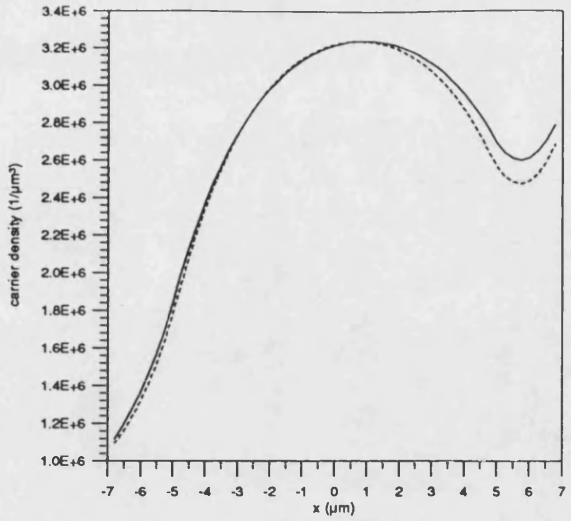


Fig. 6.3.2: Linear diffusion equation solved in a finite region ($D = 3.5\ 10^9$ μm²/s, τ = 8ns, d=0.1μm, $J_o = 3.0\ 10^{-5}$ A/μm², contact width = 10μm, $x_o = -7$μm, $N_o = 3.0\ 10^6 \mu m^{-3}$ and x_3=7μm, $N_1 = 1.0\ 10^6\ \mu m^{-3}$): solid line: HGCM (M = 99, w_o = 1.05μm), broken line: analytic solution.

3.1.3 Finite region with mixed boundary conditions

This case is the most complicated of the three. The mixed boundary conditions considered here are

$$D\frac{dN}{dx}\bigg|_{x=x_q} = -v_q\left(N(x_q) - N_q\right) \tag{6.3.6}$$

where v_q (in ms^{-1}) is the (constant) surface recombination rate, N_q (in m^{-3}) is a constant parameter, and q = 0, 3. It can be recognised that boundary conditions of the type (6.3.6) are typically associated with surface recombination processes, [6].

In this case equation (6.3.1) may be conveniently solved with the HGCM if the physical domain is embedded in an unbounded open region, as shown in Fig. 6.3.3. Complications may be encountered due to the possibility of the derivative of N(x) being discontinuous (at material interfaces) depending on the specified boundary conditions, in which case the solution would not be appropriately described by the HG functions. Thus, the corresponding consistently posed problem of solving the diffusion equation in an unbounded region is examined and hence the boundary conditions (6.3.2) need to be considered. In the fictitious unbounded region, the two lateral regions in Fig. 6.3.3 are defined by the parameters D_1, τ_1 and D_3, τ_3, respectively, so that the solution and its derivative are continuous over the infinite range.

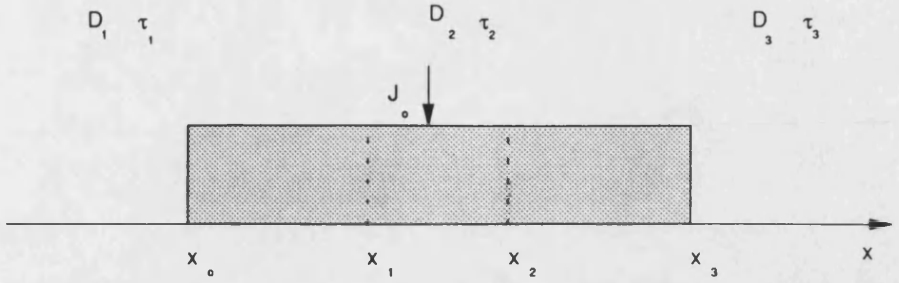


Fig. 6.3.3: Structure used for the solution of the linear diffusion equation with mixed boundary conditions (case 3.1.3). The current injection is limited to the region (x_1, x_2). The region $x < x_o$ is characterised by the parameters D_1, τ_1; the region $x_o < x < x_3$ by D_2, τ_2; and the region $x > x_3$ by D_3, τ_3.

The analytic solution of (6.3.1) in this case takes the following form:

$$N(x) = \begin{cases} Ae^{\frac{x}{\sqrt{D_1\tau_1}}} + Be^{\frac{-x}{\sqrt{D_3\tau_3}}} & x < x_o \\ Ce^{\frac{x}{\sqrt{D_2\tau_2}}} + De^{\frac{-x}{\sqrt{D_2\tau_2}}} & x_o < x < x_1 \\ Ee^{\frac{x}{\sqrt{D_2\tau_2}}} + Fe^{\frac{-x}{\sqrt{D_2\tau_2}}} + \dfrac{J_o}{qdD_2} & x_1 < x < x_2 \\ Ge^{\frac{x}{\sqrt{D_2\tau_2}}} + He^{\frac{-x}{\sqrt{D_2\tau_2}}} & x_2 < x < x_3 \\ Ie^{\frac{x}{\sqrt{D_3\tau_3}}} + Je^{\frac{-x}{\sqrt{D_3\tau_3}}} & x > x_3 \end{cases} \qquad (6.3.7)$$

with A, B, C, D, E, F, G, H, I, J, $(D_1\tau_1)$, $(D_3\tau_3)$ determined using the boundary conditions (6.3.2), the condition imposed by (6.3.6) at $x = x_o$, x_3, and the continuity of $N(x)$ and $\dfrac{dN(x)}{dx}$ at $x = x_o$, x_1, x_2, x_3, [Appendix 6.3]. The very satisfactory comparison between the HGCM and the analytic solutions is shown in Fig. 6.3.4.

In all the cases investigated in this section it has been found that the solutions are sensitive to the choice w_o. However, it has not been possible to establish any general criteria for evaluating the optimum value for w_o.

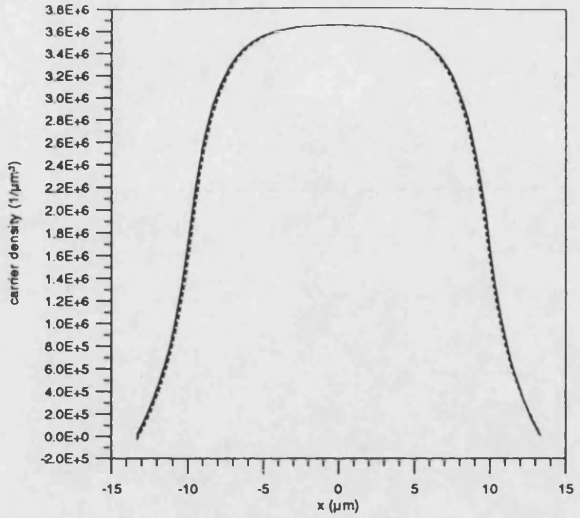


Fig. 6.3.4: Linear diffusion equation solved in a finite region with mixed boundary conditions (D_2= 3.5 10^9 µm²/s, τ = 8ns, d = 0.1µm, J_o = 3.0 10^{-5} A/µm², contact = 19.5µm, v_o = 1.0 10^{10}µm/s, $x_o = x_3 = -13.5$µm, $N_o = N_3 = 0$µm^{-3}): solid line: HGCM (M = 99, w_o = 1.0µm), broken line: analytic solution.

3.2 Nonlinear diffusion equation - comparison with other methods of solution

In this section the HGCM is used to solve the nonlinear diffusion equation of the form of equation (6.1.5). The results from the HGCM are compared with those obtained with the JTM, [1], (in that reference the JTM is compared with two other numerical methods, namely the Perturbation Method and the Transfer-Matrix-Method).

First consider an example of (active) device of the type shown in Fig. 6.1.1, in which the Auger recombination is negligible, i.e., $\gamma = 0$ in equation (6.1.5). The carrier profiles at the narrow and at the wide end of the device, calculated with the HGCM and the JTM, are compared in Fig. 6.3.5. Results from the two methods are in excellent agreement.

Equation (6.1.5) is now solved including the Auger recombination term ($\gamma > 0$). Fig. 6.3.6 shows the results calculated by the HGCM as compared with those obtained using the JTM. As in the previous case the results from the two different methods are in very good agreement.

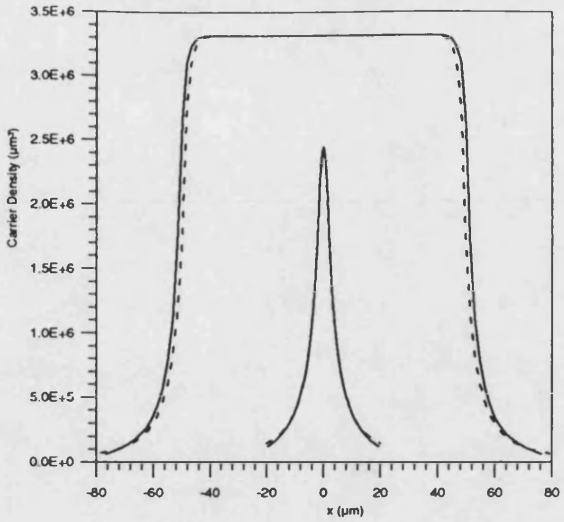


Fig. 6.3.5: Solution of the nonlinear diffusion equation: solid line: HGCM (M = 99, w_o = 1.5µm at narrow end, w_o = 6.0µm at broad end), broken line: JTM. (Contact width at narrow end = 4µm, at wide end = 100µm, D = 1.5 10^9 µm²/s, B_r = 10^2µm³/s, n_o = 10^4 µm^{-3}, J_o = 3.0 10^{-5} A/µm², d = 0.17µm.)

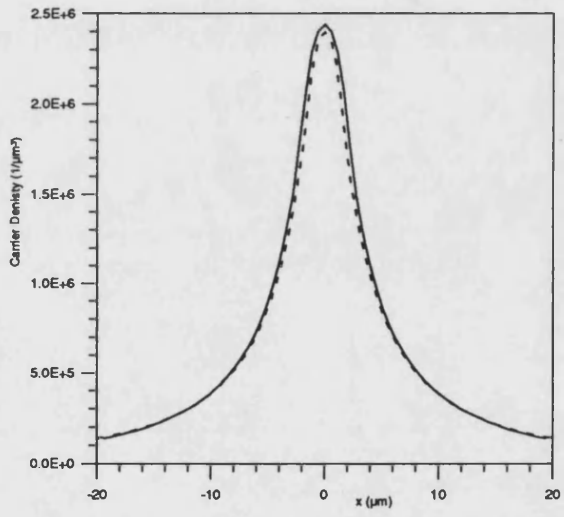


Fig. 6.3.6: Carrier profile calculated including the effect of Auger recombination (γ = 1.0 10^{-5} µm^{6}s^{-1}, other parameters as in Fig. 6.3.5, at narrow end): solid line: HGCM (M = 99, w_o = 1.5µm); broken line: JTM.

It is observed that the inclusion of the Auger recombination term increases the computational efforts required to find the self-consistent solution; in some cases (large γ) it may reduce the stability of the algorithm so that convergence may not be achieved, which is also noted with other numerical methods. However, it has been found that by writing

$$\hat{N}_k^2(x) = \tfrac{1}{2}\left(\hat{N}_{k-2}^2(x) + \hat{N}_{k-1}^2(x)\right) \qquad (6.3.8)$$

in equation (6.1.5) for the k-th iteration, where $\hat{N}_{k-2}^2(x)$ and $\hat{N}_{k-1}^2(x)$ refer to the solutions obtained in the iteration k–2 and k–1, respectively, the HGCM iteration procedure converges faster to the required solution.

Since computing time is an important consideration, Tab. 6.3.1 tabulates the computation time taken by the HGCM and by other numerical methods, [1], for the solution of the carrier diffusion equation of the example considered in Fig. 6.3.5 and in Fig. 6.3.6 (with and without the Auger recombination term). It is seen that the HGCM compares very favourably with other methods of solution, Tab. 6.3.1, even if it appears to be marginally slower than the JTM.

Tab. 6.3.1: Summary of the CPU time (in seconds) needed by the HGCM and other methods of solution - the JTM, the perturbation method (PM), and the transfer matrix method (TMM):

	HGCM		JTM		PM		TMM	
M	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
31	0.4	1.9	0.7	0.8	26.0	27.0	14.5	-
61	1.3	2.4						
99	2.3	2.6						

(1) without Auger recombination, (2) with Auger recombination

4. Sensitivity of the solution to M and w_o

The accuracy of the proposed HGCM solution scheme is controlled by two parameters: the first is the number of expansion terms (M) used in equation (6.1.3), which also corresponds to the number of collocation points; and the second is the width parameter, w_o, defining the HG functions. The extent of the computational window is determined by both parameters. Keeping the number (M) of collocation points fixed, the w_o parameter determines the extent of the transverse (x) range by altering the separation between the points; on the other hand, given a value for w_o, the number (M) of points determines the density of collocation points in that range.

The effect of w_o is investigated by solving equation (6.1.5) with different w_o values, but keeping the number of expansion terms constant (M = 99). The results are shown in Fig. 6.4.1. The solutions are qualitatively unaffected by w_o; however if w_o is too small the computational range is not sufficient for describing the solution. From Fig. 6.4.1 it can also be noticed that the peak value of N(x) varies with w_o, as shown

in detail in Fig. 6.4.2, but the fluctuations are small - the maximum relative error in Fig. 6.4.2 is, in fact, less than 5% for a wide range of w_o values.

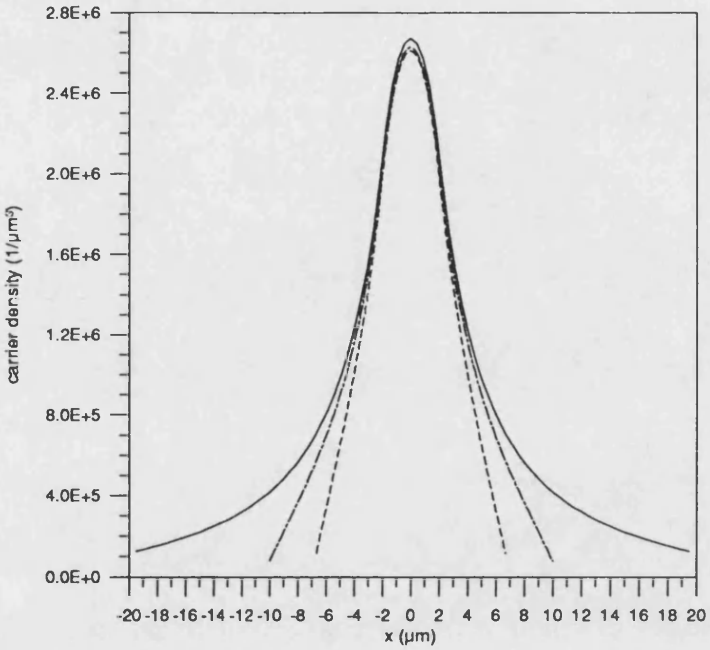


Fig. 6.4.1: Effect of w_o parameter on the solution of the diffusion equation with the HGCM (M = 99): broken line: w_o = 0.5μm; broken-dotted line: w_o = 0.75μm; solid line: w_o = 2.0μm (contact width = 4μm, D = 1.5 10^9 μm²/s, B_r = 10^2 μm³/s, n_o = 10^4 μm^{-3}, J_o = 3.0 10^{-5} A/μm², d = 0.17 μm).

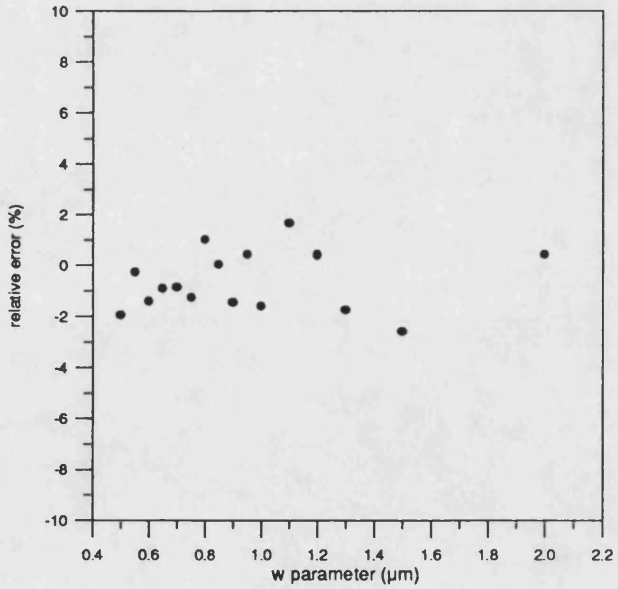


Fig. 6.4.2: Peak value of the carrier distribution as a function of the width parameter used in the HGCM (M = 99). The variation is less than ±2%, (contact width = 4μm, D = 1.5 10^9 μm²/s, B_r = 10^2 μm³/s, n_o = 10^4 μm^{-3}, J_o = 3.0 10^{-5} A/μm², d = 0.17 μm).

The effect of the number of collocation points, M, on the solution of the carrier diffusion equation is next investigated. The carrier profiles shown in Fig. 6.4.3 have been calculated using three different values of M, but at the same time choosing the w_o value such that the x-range is unchanged. As expected, the accuracy of the solution determined with the HGCM increases with the number of expansion terms M. Also, as in the previous case, it is noted that the peak value of N(x) fluctuates with M; however, the fluctuations are of 5% relative error, as shown in Fig. 6.4.4.

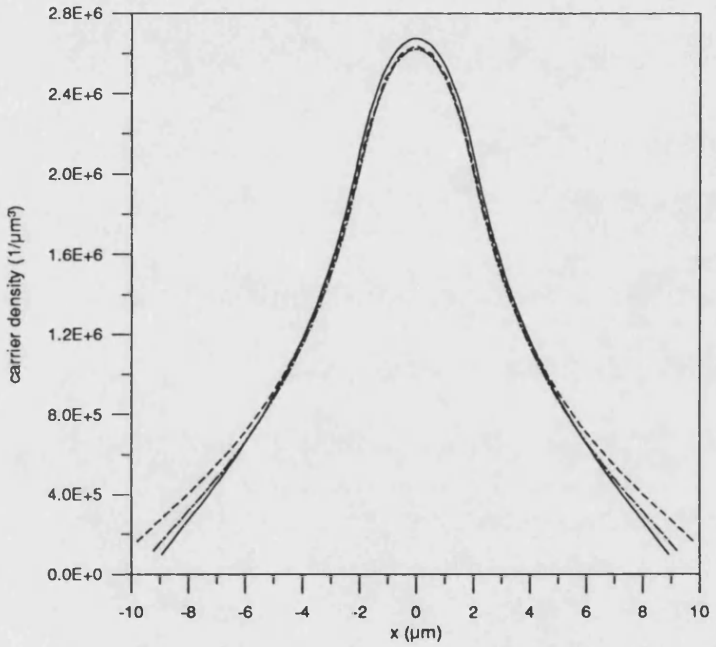


Fig. 6.4.3: Carrier profile calculated with the HGCM with varying M: broken line: M = 31; broken-dotted line: M = 61; solid line: M = 111 (contact width = 4μm, D = 1.5 10^9 μm²/s, B_r = 10^2 μm³/s, n_o = 10^4 μm^{-3}, J_o = 3.0 10^{-5} A/μm², d = 0.17 μm).

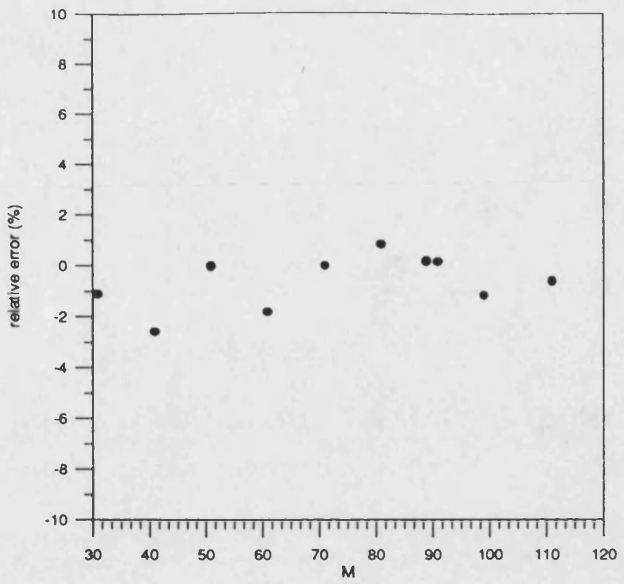


Fig. 6.4.4: Dependence of the peak value of the carrier distribution profile with varying M. Note that the variation is less than ±2% (contact width = 4μm, D = 1.5 10^9 $\mu m^2/s$, B_r = 10^2 $\mu m^3/s$, n_o = 10^4 μm^{-3}, J_o = 3.0 10^{-5} $A/\mu m^2$, d = 0.17 μm).

However, a larger number of points also implies a larger computational effort. From Tab. 6.3.1 it is observed that the time needed for computation increases approximately as M^2 and hence M should be chosen wisely by compromising between accuracy and CPU time.

From the above discussions it emerges that the solution of the nonlinear diffusion equation (6.1.5) is relatively insensitive to the two parameters, w_o and M, which may be qualitatively explained by the fact that the iterative scheme converges for the range of (device) parameters used in the calculations. This implies that in the self-consistent model described in the next chapter, the two parameters w_o and M can be chosen according to the requirements specified by the field equation.

Finally, as discussed in Section 1 of this chapter, since the effects of diffusion on the carriers profile varies according to the width of the contact of current injection the initial profile can be appropriately chosen according to the contact width, to be either a Gaussian or a top-hat distribution, in order to optimise on the iteration process. Although this process reduces the CPU time needed by the HGCM, it is noted that the HGCM converges to the required solution of the diffusion equation, independent of the initial carrier distribution, $N_o(x)$ in Fig. 6.2.1.

Summary

In this chapter the HGCM has been used to solve the carrier diffusion equation typically associated with (active) semiconductor optical devices. The HGCM formalism that has been used in connection with the wave equation can be used almost unchanged in the context of the diffusion equation.

Linear forms of the diffusion equation have been solved with the proposed HGCM scheme, and very good agreement with the analytic solutions has been achieved. Results calculated with the HGCM for the nonlinear diffusion equation have been compared with those from another method of solution, namely the Jacobi-Tridiagonal method - again, excellent agreement is observed. It is noted that the HGCM solution of the nonlinear diffusion equation is relatively insensitive to the choice of the w_o parameter, which is qualitatively explained by the consideration that the nonlinear equation converges for the range of parameters used in the analysis.

Appendix 6.1

The analytic procedure to solve the carrier diffusion equation with the HG expansion method

Substituting the function expansion (6.1.3) into the diffusion equation (6.1.5) obtain the following equation:

$$D\sum_{k=1}^{\infty} a_k f_k''(x) - \left\{B_r\left[n_o + \hat{N}\right] - \gamma\hat{N}^2\right\}\sum_{k=1}^{\infty} a_k f_k(x) + \frac{J(x)}{qd} = 0 \quad \text{(A6.1 - 1)}$$

[for convenience $w_o = 1$ has been used]. Then, using equation (2.2.6), and orthogonalising obtain

$$-D(2j+1)a_j + D\sum_{k=1}^{\infty} a_k \int_{-\infty}^{+\infty} f_j^*(x)\left[x^2 - \left\{B_r\left[n_o + \hat{N}\right] - \gamma\hat{N}^2\right\}\right] f_k(x)dx + \\ + \frac{1}{qd}\int_{-\infty}^{+\infty} f_j^*(x)J_o(x)dx = 0 \quad \text{(A6.1 - 2)}$$

The integrals in equation (A6.1 - 2) can be discretised using the Gaussian Quadrature formula (A4.3 - 2), [5], [Appendix 4.3]

$$\int_{-\infty}^{+\infty} F(x)dx \approx \sum_{i=1}^{M} F(x_i)e^{x_i^2} h_i \quad \text{[A4.3 - 2]}$$

for any function F(x) in the appropriate functional space, where h_i are the weight functions for the Hermite polynomials, $h_i = \frac{2^{m+1} m!\sqrt{\pi}}{\left[H_m'(x_i)\right]^2}$, x_i the sampling points corresponding to the (M) roots of the Hermite polynomial $H_M(x)$.

Using (A4.3 - 2), equation (A6.1 - 2) is transformed into the following matrix equation

$$D\underline{\underline{G}}^T\underline{\underline{H}}\underline{\underline{C}}_2\underline{\underline{G}}\underline{\underline{A}} - D\underline{\underline{C}}_1\underline{\underline{A}} - \underline{\underline{G}}^T\underline{\underline{H}}\left[B_r\left(n_o\underline{\underline{I}} + \underline{\underline{N}}_1\right) + \gamma\underline{\underline{N}}_2\right]\underline{\underline{G}}\underline{\underline{A}} + \underline{\underline{G}}^T\underline{\underline{H}}\underline{J} = 0 \quad \text{(A6.1 - 3)}$$

where $\underline{\underline{G}}$ is the matrix of the HG functions evaluated at the sampling points, defined by $\underline{\underline{G}} = \left\{f_j(x_i); i, j = 1, 2, \ldots, M\right\}$;

$\underline{\underline{G}}^T$ is the transpose of $\underline{\underline{G}}$;

$\underline{\underline{H}}$ the diagonal matrix of the weight functions, $\underline{\underline{H}} = \left\{\text{diag}\left(e^{x_i^2} h_i\right); i = 1, 2, \ldots, M\right\}$;

$\underline{A}$ the vector of the (constant) unknown expansion coefficients, $\underline{A} = [a_i; i = 1, 2, \ldots, M]$;

$\underline{\underline{C}}_1$ the diagonal matrix defined by $\underline{\underline{C}}_1 = \{\text{diag}(2i-1); i = 1, 2, \ldots, M\}$;

$\underline{\underline{C}}_2$ the diagonal matrix defined by $\underline{\underline{C}}_2 = \left\{\text{diag}\left(x_i^2\right); i = 1, 2, \ldots, M\right\}$;

$\underline{\underline{N}}_1$ the diagonal matrix of the values of the carrier distribution sampled at the collocation points calculated in the previous iteration (or the initial guess at the beginning of the iteration process), $\underline{\underline{N}}_1 = \left\{\text{diag}\left(\hat{N}(x_i)\right); i = 1, 2, \ldots, M\right\}$;

$\underline{\underline{N}}_2$ the diagonal matrix defined by $\underline{\underline{N}}_2 = \left\{\text{diag}\left(\hat{N}^2(x_i)\right); i = 1, 2, \ldots, M\right\}$;

$\underline{\underline{I}}$ the identity matrix;

$\underline{J}$ the vector of the current density profile sampled at the collocation points $\underline{J} = \left[\frac{J(x_i)}{qd}; i = 1, 2, \ldots, M\right]$;

and the other symbols are as defined in Section 2.

Since the HG functions are orthogonal and the sampling points are all distinct and real, the inverse of matrix $\underline{\underline{G}}$ always exists, [Appendix 4.5], and can be calculated by

$$\underline{\underline{G}}^{-1} = \underline{\underline{G}}^T \underline{\underline{H}} \qquad \text{[A4.5 - 11]}$$

Using (A6.1 - 4) in (A6.1 - 3) obtain

$$\underline{\underline{P}}\underline{A} + \underline{\underline{G}}^{-1}\underline{J} = 0 \qquad \text{(A6.1 - 4)}$$

where $\underline{\underline{P}} = D\left(\underline{\underline{G}}^{-1}\underline{\underline{C}}_2\underline{\underline{G}} - \underline{\underline{C}}_1\right) - B_r\underline{\underline{G}}^{-1}\left(n_o\underline{\underline{I}} + \underline{\underline{N}}_1\right)\underline{\underline{G}} - \gamma\underline{\underline{G}}^{-1}\underline{\underline{N}}_2\underline{\underline{G}}$ is an invertible matrix. Hence, in principle, the expansion coefficients used in (6.1.3) can be calculated solving (A6.1 - 4).

Alternatively (A6.1 - 4) may be written in terms of $\underline{N}$ using the identity $\underline{N} = \underline{\underline{G}}\underline{A}$, and hence

$$\underline{\underline{T}}\underline{N} + \underline{J} = 0 \qquad \text{(A6.1 - 5)}$$

with $\underline{\underline{T}} = D\left(\underline{\underline{C}}_2 - \underline{\underline{G}}\underline{\underline{C}}_1\underline{\underline{G}}^{-1}\right) - B_r\left(n_o\underline{\underline{I}} + \underline{\underline{N}}_1\right) - \gamma\underline{\underline{N}}_2$.

Equation (A6.1 - 5) can also be obtained by applying the collocation method directly to the diffusion equation, as presented in Appendix 6.2, establishing the correspondence between the orthogonalisation process described above and the collocation method.

Appendix 6.2

The collocation method for solving the carrier diffusion equation

Substituting the HG expansion (6.1.3) into the diffusion equation (6.1.5), and writing the resulting differential equation at each collocation point obtain

$$D\sum_{k=1}^{M} a_k f_k''(x_i) - \left\{B_r\left[n_o + \hat{N}(x_i)\right] + \gamma\hat{N}^2(x_i)\right\}N(x_i) + \frac{J(x_i)}{qd} = 0 \qquad \text{(A6.2 - 1)}$$

Then making use of the Sturm-Liouville equation (2.2.6) for the HG functions [Chapter 2], equation (A6.2 - 1) reduces to

$$D\sum_{k=1}^{M} a_k\left\{(x_i^2 - 2k - 1) - \left[B_r\left[n_o + \hat{N}(x_i)\right] + \gamma\hat{N}^2(x_i)\right]\right\}f_k(x_i) + \frac{J(x_i)}{qd} = 0 \qquad \text{(A6.2 - 2)}$$

Equation (A6.2 - 2) may be conveniently written in the compact matrix form

$$D\left(\underline{\underline{C}}_2\,\underline{\underline{G}} - \underline{\underline{G}}\,\underline{\underline{C}}_1\right)\underline{A} - B_r\left(n_o\underline{\underline{I}} + \underline{\underline{N}}_1\right)\underline{N} - \gamma\underline{\underline{N}}_2\underline{N} + \underline{J} = 0 \qquad \text{(A6.2 - 3)}$$

where $\underline{\underline{G}}$, $\underline{A}$, $\underline{\underline{C}}_1$, $\underline{\underline{C}}_2$, $\underline{\underline{N}}_1$, $\underline{\underline{N}}_2$, $\underline{\underline{I}}$, $\underline{J}$ are as defined in Appendix 6.1, and the other symbols are as defined in Section 2.

From equation (6.1.3) it follows that the (sampled) carrier profile can be written in the form of a matrix-vector product

$$\underline{N} = \underline{\underline{G}}\,\underline{A} \qquad \text{(A6.2 - 4)}$$

Hence, since the matrix $\underline{\underline{G}}$ can always be inverted [Appendix 4.5]

$$\underline{A} = \underline{\underline{G}}^{-1}\underline{N} \qquad \text{(A6.2 - 5)}$$

so that equation (A6.2 - 3) reduces to the following matrix equation in $\underline{N}$:

$$\underline{\underline{T}}\,\underline{N} + \underline{J} = 0 \qquad \text{(A6.2 - 6)}$$

where $\underline{\underline{T}} = D\left(\underline{\underline{C}}_2 - \underline{\underline{G}}\,\underline{\underline{C}}_1\,\underline{\underline{G}}^{-1}\right) - B_r\left(\underline{\underline{N}}_1 + n_o\underline{\underline{I}}\right) - \gamma\underline{\underline{N}}_2$.

Appendix 6.3

Linear diffusion equation: analytic solution

1. Unbounded region

The coefficients for the analytic solution

$$N(x)=\begin{cases} B_1 e^{\frac{x}{\sqrt{D\tau}}} & x < x_1 \\ C_1 e^{\frac{-x}{\sqrt{D\tau}}} + C_2 e^{\frac{+x}{\sqrt{D\tau}}} + \dfrac{J_o \tau}{qd} & x_1 < x < x_2 \\ D_1 e^{\frac{-x}{\sqrt{D\tau}}} & x > x_2 \end{cases} \qquad [6.3.3]$$

are $\quad B_1 = \dfrac{b}{2a}\left(e^{-x_1\sqrt{a}} - e^{-x_2\sqrt{a}}\right), \quad C_1 = \dfrac{-b}{2a} e^{x_1\sqrt{a}}, \quad C_2 = \dfrac{-b}{2a} e^{-x_2\sqrt{a}},$

$D_2 = \dfrac{b}{2a}\left(e^{x_2\sqrt{a}} - e^{x_1\sqrt{a}}\right)$ with $a = \dfrac{1}{D\tau}$ and $b = \dfrac{J_o}{Dqd}$.

2. Finite region with uniform boundary conditions

The coefficients for the analytic solution

$$N(x)=\begin{cases} B_1 e^{\frac{-x}{\sqrt{D\tau}}} + B_2 e^{\frac{x}{\sqrt{D\tau}}} & x_o < x < x_1 \\ C_1 e^{\frac{-x}{\sqrt{D\tau}}} + C_2 e^{\frac{+x}{\sqrt{D\tau}}} + \dfrac{J_o \tau}{qd} & x_1 < x < x_2 \\ D_1 e^{\frac{-x}{\sqrt{D\tau}}} + D_2 e^{\frac{x}{\sqrt{D\tau}}} & x_2 < x < x_3 \end{cases} \qquad [6.3.5]$$

are $\quad C_1 = \dfrac{N_o e^{\sqrt{a}(x_3-x_o)} - N_1 + \frac{b}{a}Ch\left(\sqrt{a}[x_2-x_3]\right) - \frac{b}{a}Ch\left(\sqrt{a}[x_o-x_1]\right)e^{\sqrt{a}(x_2-x_o)}}{2Sh\left(x_3\sqrt{a}\right)},$

$C_2 = \dfrac{N_1 - N_o e^{\sqrt{a}(-x_o-x_3)} - \frac{b}{a}Ch\left(\sqrt{a}[x_2-x_3]\right) + \frac{b}{a}e^{\sqrt{a}(-x_o-x_3)}Ch\left(\sqrt{a}[x_o-x_1]\right)}{2Sh\left(x_3\sqrt{a}\right)},$

$B_1 = C_1 + \dfrac{b}{2a}e^{+x_1\sqrt{a}}, \; B_2 = C_2 + \dfrac{b}{2a}e^{-x_1\sqrt{a}},$

$D_1 = C_1 + \dfrac{b}{2a}e^{x_2\sqrt{a}}, \; D_2 = C_2 + \dfrac{b}{2a}e^{-x_2\sqrt{a}}$

with $a = \dfrac{1}{D\tau}$ and $b = \dfrac{J_o}{Dqd}$.

c) finite region with mixed boundary conditions

The coefficients for the analytic solution

$$N(x)=\begin{cases} Ae^{\frac{x}{\sqrt{D_1\tau_1}}}+Be^{\frac{-x}{\sqrt{D_3\tau_3}}} & x<x_o \\ Ce^{\frac{x}{\sqrt{D_2\tau_2}}}+De^{\frac{-x}{\sqrt{D_2\tau_2}}} & x_o<x<x_1 \\ Ee^{\frac{x}{\sqrt{D_2\tau_2}}}+Fe^{\frac{-x}{\sqrt{D_2\tau_2}}}+\dfrac{J_o}{qdD_2} & x_1<x<x_2 \\ Ge^{\frac{x}{\sqrt{D_2\tau_2}}}+He^{\frac{-x}{\sqrt{D_2\tau_2}}} & x_2<x<x_3 \\ Ie^{\frac{x}{\sqrt{D_3\tau_3}}}+Je^{\frac{-x}{\sqrt{D_3\tau_3}}} & x>x_3 \end{cases} \qquad [6.3.7]$$

are $B = 0, I = 0,$

$$F=\frac{b}{2}\frac{e^{(x_2-x_3)\sqrt{a_2}}-\dfrac{\sqrt{a_2}+\sqrt{a_3}}{\sqrt{a_2}-\sqrt{a_3}}\left(e^{(x_3-x_1)\sqrt{a_2}}-e^{-x_1\sqrt{a_2}}+\dfrac{\sqrt{a_2}+\sqrt{a_1}}{\sqrt{a_2}-\sqrt{a_1}}e^{(x_1-2x_o)\sqrt{a_2}}\right)}{\dfrac{\sqrt{a_2}+\sqrt{a_1}}{\sqrt{a_2}-\sqrt{a_1}}\dfrac{\sqrt{a_2}+\sqrt{a_3}}{\sqrt{a_2}-\sqrt{a_3}}e^{-x_o\sqrt{a_2}}-e^{-x_3\sqrt{a_3}}},$$

$$E=\frac{\sqrt{a_2}+\sqrt{a_1}}{\sqrt{a_2}-\sqrt{a_1}}e^{-x_o\sqrt{a_2}}F+\frac{b}{2}\frac{\sqrt{a_2}+\sqrt{a_1}}{\sqrt{a_2}-\sqrt{a_1}}e^{(x_1-2x_o)\sqrt{a_2}}-\frac{b}{2}e^{-x_1\sqrt{a_2}},$$

$$C=E+\frac{b}{2a_2}e^{-x_1\sqrt{a_2}},\quad D=F+\frac{b}{2a_2}e^{x_1\sqrt{a_2}},$$

$$G=E+\frac{b}{2a_2}e^{-x_2\sqrt{a_2}},\quad H=F+\frac{b}{2a_2}e^{x_2\sqrt{a_2}},$$

$$A=Ce^{x_o(\sqrt{a_1}-\sqrt{a_1})}+De^{-x_o(\sqrt{a_2}-\sqrt{a_1})},$$

$$J=Ge^{x_3(\sqrt{a_2}+\sqrt{a_3})}+He^{-x_3(\sqrt{a_2}-\sqrt{a_3})},$$

$$\sqrt{a_1}=-\frac{x_o}{|x_o|}\frac{v_o}{D},\quad \sqrt{a_3}=\frac{x_3}{|x_3|}\frac{v_1}{D}.$$

For brevity the following notation has been used: $a_k=\dfrac{1}{D_k\tau_k}$ with k = 1, 2, 3, and

$$b_2=\frac{J_o}{D_2qd}.$$

References

[1] P. S. Spencer, R. Balasubramanyan, J. Sarma, K. A. Shore, *Self-Consistent Solution of the Diffusion Equation for an Active Optical Semiconductor Device*, Semiconductor Science and Technology, 10 (1995), pp. 942-947

[2] G. H. B. Thompson, *Physics of semiconductor laser devices*, John Wiley (1980)

[3] G. P. Aggrawal and N. K. Dutta, *Long-Wavelength Semiconductor Lasers*, Van Nostrand Reinhold (1986)

[4] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill Book Company (1953)

[5] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications (1965)

[6] S. M. Sze, *Physics of semiconductor devices*, John Wiley (1981)

Chapter 7

The HGCM for the self-consistent solution of active semiconductor optical devices

Two computational schemes based on the HGCM have been developed in the previous chapters. The first has been used for the analysis of electromagnetic field propagation in (longitudinally non-uniform) optical structures. The second scheme, basically equivalent to the first, solves the carrier diffusion equation. In the present chapter these two HGCM schemes will be combined to derive a comprehensive model for the analysis of active optical devices.

A particular feature of active semiconductor optical devices is that the injected carriers (inversion population) induces changes in the refractive index distribution which, in turn, affect the optical field in the cavity via the diffusion equation. In terms of modelling, the field-carrier interaction can be accounted for by coupling the wave equation to the carrier diffusion equation through the stimulated recombination term. Ideally the two equations should be solved simultaneously to correctly describe the operation of active devices. However, the solution is typically achieved by an iteration procedure ensuring that the two equations are solved self-consistently. Particularly for the modelling of flared active devices the inherent longitudinal non-uniformity of the structure introduces additional complications especially in the analysis of the field distribution since analytic solutions are generally not possible.

This chapter is structured as follows: in Section 1 the background history of tapered devices is summarised. The proposed self-consistent scheme based on the HGCM is presented in Section 2. The first set of results is discussed in Section 3 and refers to linearly tapered active devices. In particular in Section 3.1 travelling-wave amplifiers are modelled and the results obatined with the HGCM are compared with those calculated with another self-consistent model. In Section 3.2 the self-consistent analysis of laser devices is discussed, showing various stages of self-consistency obtained with the HGCM. Parabolic tapered lasers are thus modelled in Section 4, where results from the HGCM are compared with those obtained with another method of solution. Two examples of index- and gain-guided stripe lasers are

considered in Section 5 as a further test of the modelling capabilities of the HGCM. Finally, in Section 6 two alternative self-consistent methods, one based on the BPM, the other on the mode matching method, are reviewed in the context of active devices to highlight the different features characterising those two models with respect to the HGCM presented in the previous sections.

Note that the results discussed in this chapter derive from a steady state analysis of active optical devices. Time dependent effects may significantly affect the operation of semiconductor active devices in some circumstances (e.g., filamentation, chaotic behaviour) and strictly they should be included, [1], [2]. The incorporation of time dependent effects in the HGCM model is in fact contemplated as future work, and it is suggested that the steady-state HGCM model presented here constitutes a convenient basis for the extension to a more sophisticated time and space dependent model for (tapered) active devices.

1. Tapered laser devices

The objective of achieving integrated optical devices and circuits explains the intense research effort invested in the analysis and improvement of semiconductor devices. Increasing interest has recently been directed towards the achievement of high power (semiconductor) optical sources for applications in optical pumping, laser printing, communications, etc. . The most immediate way of generating high output powers from ordinary (narrow) stripe lasers, Fig. 7.1.1 a), is to increase the injection current. However, apart from saturation effects, catastrophic optical damage may occur at the resonator facets because of the high optical power densities created in the laser cavity. To avoid such detrimental effects, arrays of ordinary narrow stripe lasers, Fig. 7.1.1 b), have been used which have shown stable optical characteristics even at high injection levels. However, the lasing mode of multi-stripe laser arrays is generally the fundamental (lowest loss) transverse eigenmode in which the individual emitters are in antiphase. This feature produces undesirable double lobed far-field patterns from multistripe lasers, [3]. For applications that require near diffraction limited beams (such as optical pumping, printing, medical imaging, sensing, ...) other solutions have to be adopted. An alternative approach is to increase the lateral dimension of ordinary stripe lasers to operate as broad area (BA) lasers, Fig. 7.1.1 c).

In fact, increasing the volume of the lasing mode results in an increase of the laser output power. However, the drawback of BA lasers, is that they tend to excite multiple lateral modes. At or near threshold BA devices may operate as near to diffraction limited sources, [3], but at higher injection currents the far field patterns produced by BA lasers are typically unstable, irregular and broadening with increasing injection currents, [3], [4], [5], [6]. Further, BA lasers are subject to filamentation because of self-focusing [4], [6], [7], which makes the devices spatially incoherent at the output facet. Various configurations have been proposed to reduce such instability problems which are intrinsic to BA lasers, e.g., incorporation of unstable resonators, tailoring of the gain region, injection-locking, thermally induced or built-in lateral index tailoring, incorporation of an external cavity, [3] and references therein. However, although most of these configurations have been demonstrated to stabilise BA lasers and to enhance the control on the lateral-mode content of the output beam even at high output power levels, [3], the additional design and fabrication requirements are generally very demanding. Recently active devices have been designed with tapered geometries, [8], [9], [10], Fig. 7.1.1 d), to reduce catastrophic optical damage, filamentation and saturation effects. Tapered semiconductor devices, in fact, seem to provide structures capable of producing most of the beam characteristics needed for applications ranging from mode-converters [11] to high-power optical sources [8], [9], [10], [12], [13]. In addition to the desirable operational features, the advantages of easy fabrication and integration justify the efforts directed towards a full understanding of the properties of flared devices.

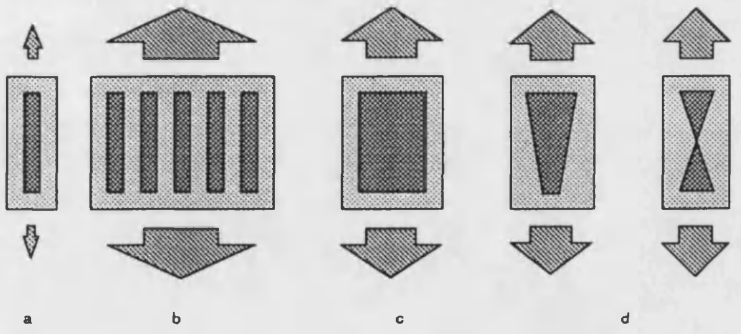


Fig. 7.1.1: Towards tapered geometry devices: a) ordinary stripe laser, b) multi-stripe laser array, c) broad area laser, d) tapered geometry laser and bow-tie laser.

2. Active optical devices modelling: the HGCM formulation

The devices of interest are assumed to be fabricated from typical double heterostructure multilayer semiconductor laser material [Appendix 7.1]. With reference to Fig. 7.2.1, the dominant optical confinement effect is along the vertical (y) axis. Along the horizontal (x) direction the devices analysed in this thesis typically present a rib structure, Fig. 7.2.1 b), which not only reduces current spreading, but also leads to (weak) optical confinement along the x-axis. Thus, the definition of weakly guiding structures for the devices considered in the present context, refers to the analysis of the guiding properties in the (x,z) plane. It is convenient to consider the strong guiding effect separately by applying the effective dielectric constant method along the y-axis [Appendix 1.1], or by any other appropriate method, to reduce the three-dimensional analysis to that in two dimensions. Thus, with reference to Fig. 7.2.1 c), the transverse (x) and the longitudinal (z) axes are of interest.

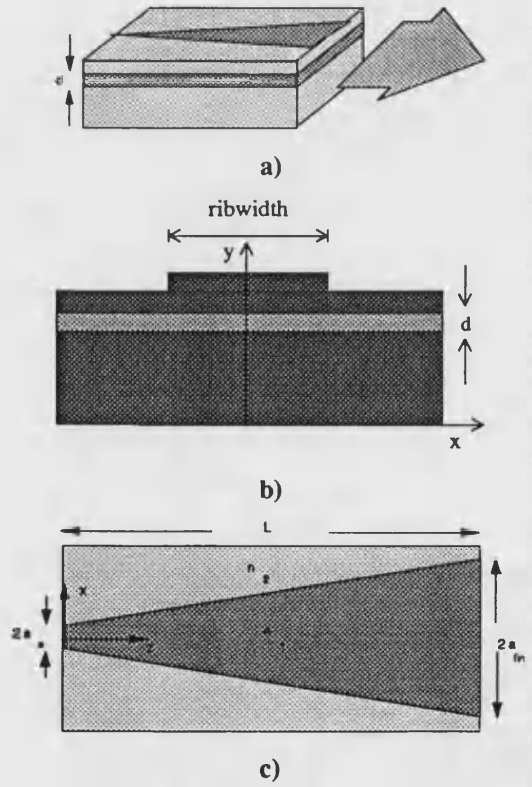


Fig. 7.2.1: a) Tapered device, b) front view: d is the thickness of the active layer, c) top-plane view: $2a_o$ and $2a_{fin}$ are the widths of narrower and wider facets, respectively; L is the length of the device. The tapered shaded region also corresponds to the area of the metal contact.

In the present formulation reference is made to the phenomenological approach, [14], typically used for the modelling of active semiconductor devices. The optical signal will be assumed to have a fixed free space wavelength, λ_o, corresponding to a frequency $\omega_o = \frac{2\pi c}{\lambda_o}$. Spontaneous emission is neglected entirely in the present formulation.

2.1 Field equation

In this thesis the HGCM is formulated as a scalar, total field analysis. Polarisation effects are thus neglected and hence the scalar field in the structure, F(x,z), may represent either the transverse component, E_x, of the electric field of the $(TE)_y$ type polarisation or the transverse magnetic component, H_x, of the $(TM)_y$ polarisation.

In addition to the nonlinearites produced by the field-carrier interaction, further complications in the model arise from the longitudinal non-uniformity of the tapered geometry devices. The formalism, however, may be simplified by considering paraxial propagation for the counter propagating fields in the cavity. Hence, as discussed in Chapter 1, Section 7, the total field in the dielectric structure is written as

$$F(x,z) = F_+(x,z) + F_-(x,z) = f_+(x,z)e^{-ipz} + f_-(x,z)e^{+ipz} \qquad (7.2.1)$$

$F_+(x, z)$ and $F_-(x, z)$ are the forward and reverse counter propagating fields, respectively, and $f_+(x, z)$ and $f_-(x, z)$ are the corresponding slowly varying envelopes, and p is a suitably chosen constant to take account of the fast phase change of the fields along z, [Chapter 1, Section 6.2 and Section 7]. In the present formulation the field equations to be solved are

$$\partial_x^2 f_+(x,z) - 2ip\partial_z f_+(x,z) + \left(k_o^2\varepsilon(x,z) - p^2\right)f_+(x,z) = 0 \qquad (7.2.2)$$

for the forward travelling wave, and

$$\partial_x^2 f_-(x,z) + 2ip\partial_z f_-(x,z) + \left(k_o^2\varepsilon(x,z) - p^2\right)f_-(x,z) = 0 \qquad (7.2.3)$$

for the reverse travelling wave. Using the paraxial wave equation in both directions, the forward and the reverse travelling waves are treated separately. The resulting approximate analysis that uses equations (7.2.2) and (7.2.3) considerably reduces the complications associated with the full analysis (7.2.1). However in the present formulation the coupling between the forward and the reverse travelling field is still

retained since account is taken for the boundary conditions imposed by the structure and for the complex refractive index distribution in the active medium which changes self-consistently with the optical field in the structure, as discussed in the following sections.

In the modelling of a laser device the facet reflectivities provide the boundary condition for the z axis,

$$F_+(x, z=0) = \sqrt{R_1}\, F_-(x, z=0) \tag{7.2.4}$$

and

$$F_-(x, z=L) = \sqrt{R_2}\, F_+(x, z=L) \tag{7.2.5}$$

where R_1 and R_2 are the facet (intensity) reflectivity coefficients of the facets of the devices extending between $0 < z < L$.

2.2 Complex refractive index distribution

In both equations (7.2.2) and (7.2.3), $\varepsilon(x, z) = n^2(x, z, N(x, z)) = (n_R + in_I)^2$ is the built-in and carrier dependent complex dielectric distribution. The explicit dependence of n_R and n_I on the carrier density and on the field intensity is discussed below, [14]. The complex refractive index in the device is assumed to be of the following form:

$$n(x, z, N) = n_b(x, z) + \Delta n_R(x, z, N) + in_I(x,z) + i\Delta n_I(x, z, N) \tag{7.2.6}$$

where $n_b(x, z)$ is the built-in refractive index of the structure, $n_I(x,z)$ represents the scattering losses and the free carriers absorption losses

$$n_I(x,z) = -\frac{\alpha_{sc} + \Gamma\alpha_a + (1-\Gamma)\alpha_c}{2k_o} \tag{7.2.7}$$

where α_{sc} is the scattering loss coefficient (in m^{-1}), α_a and α_c are the loss coefficients (in m^{-1}) due to the free carriers in active and cladding layer respectively, and Γ is the vertical confinement factor along the y-axis, Fig. 7.2.1 b).

$\Delta n_R(x, z, N)$ and $\Delta n_I(x, z, N)$ in (7.2.6) denote the changes due to the injected carriers and the optical field in the device. In particular, Δn_R accounts for changes in the (real part) refractive index due to the injected carriers:

$$\Delta n_R(x, z, N) = -\Gamma a_c N(x, z) \tag{7.2.8}$$

with a_c the constant differential parameter and $N(x, z)$ the injected carrier distribution.

The last term Δn_I of equation (7.2.6) accounts for variations in the gain due to the carriers injected in the device,

$$\Delta n_I(x,z,N) = \frac{\Gamma g(x,z,N)}{2k_o} \tag{7.2.9}$$

where g(x, z, N) is the material (local) gain (in m^{-1}) in the active layer of the device. In the literature the changes in the (complex) refractive index due to the injected carriers are often given in the following compact form, which is equivalent to equations (7.2.7) and (7.2.9),

$$\Delta n = \Delta n_R + i\Delta n_I = \Gamma(-\beta_a + i)\frac{g(N)}{2k_o} \tag{7.2.10}$$

where β_a is known as the antiguiding factor, [14].

In the present model saturation effects (gain nonlinearity/gain compression, [15]) are included by using the following expression for the material gain

$$g(x,z,N) = \frac{g_u(x,z,N)}{1+\frac{S_{av}(x,z)}{S_{sat}}} \tag{7.2.11}$$

with the unsaturated local gain, g_u, is assumed to be linearly dependent on the carrier density, i.e.,

$$g_u(x, z, N) = \alpha(N(x, z) - N_{tr}) \tag{7.2.12}$$

α is the gain coefficient (in m^2), N_{tr} the carrier density needed to achieve transparency (in m^{-3}); S_{sat} is the saturation power density (in Wm^{-2}), and $S_{av}(x,z) \propto \left(|f_+(x,z)|^2 + |f_-(x,z)|^2\right)$ the optical power density (in Wm^{-2}) weighted averaged across the vertical (y) direction, [Appendix 7.2].

2.3 Carrier diffusion equation

The phenomenological approach used in the previous section to establish the dependence of the refractive index on the carriers in the cavity, is completed by relating the carrier density to the pump parameter, which is the injection current density, J, [14]. This is achieved by including in the carrier diffusion equation all the (radiative and non radiative) recombination processes that modify the carrier concentration.

The devices of interest are assumed to be slowly varying along the z-axis, so that diffusion along the longitudinal (z) direction may be neglected. Hence the (time-independent) nonlinear diffusion equation, [Chapter 6, Section 1], becomes

$$D\frac{\partial^2 N(x,z)}{\partial x^2} - B_r[N(x,z)+n_o]N(x,z) - \gamma N^3(x,z) - g(x,z)S_{av}(x,z) + \frac{J(x,z)}{qd} = 0 \tag{7.2.13}$$

where D is the diffusion coefficient (in m^2s^{-1}); B_r the spontaneous recombination constant (in m^3s^{-1}); γ the Auger recombination coefficient(in m^6s^{-1}); g(x, z) the local gain as defined in equation (7.2.10); $S_{av}(x, z)$ the photon density averaged across the active layer (in $m^{-2}s^{-1}$) [Appendix 7.2]; n_o the doping density (in m^{-3}); J_o the injection current density (in Am^{-2}); d the active layer thickness (in m); q the magnitude of the electronic charge (in C).

Considering z as a parameter and the linearisation process discussed in Chapter 6, Section 1, equation (7.2.13) may be rewritten as

$$D\frac{d^2 N(x;z)}{dx^2} - B_r[\hat{N}(x;z)+n_o]N(x;z) - \gamma\hat{N}^2(x;z)N(x;z) - g(x;z)S_{av}(x;z) + \frac{J(x;z)}{qd} = 0 \tag{7.2.14}$$

where $\hat{N}(x,z)$ represents the carrier profile obtained in the previous iteration.

The boundary conditions associated with equation (7.2.14) follow from the assumption of infinite lateral extent of the device, i.e., at any z

$$N(x;z) \to 0 \quad \text{and} \quad \frac{dN(x;z)}{dx} \to 0 \qquad \text{for } |x| \to \infty \tag{7.2.15}$$

The iteration scheme used to self-consistently solve the system of coupled equations for the field and the carrier distribution is illustrated in the next section.

2.4 Self-consistent method of solution

The success of the HGCM to solve the differential equations for the field, (7.2.2) and (7.2.3), and the carrier distribution, (7.2.14), have been presented in Chapters 4 and 6. In the present discussion, the two computational schemes are combined in an iteration procedure to obtain a self-consistent numerical method that models active semiconductor optical devices.

The solutions of interest, $f_+(x,z)$, $f_-(x,z)$ and $N(x,z)$, are square integrable functions, and hence they may be individually represented in terms of a suitable (complete) set of basis functions. Since the overall self-consistent model is based on the HGCM, it is convenient to use the same HG basis set to expand both the field and the carrier profile, as discussed in Chapter 4 and Chapter 6. Hence, the forward (+) and reverse (–) travelling fields are expressed as

$$f_{\pm}(x,z) = \sum_{m=0}^{M} a_m^{\pm}(z) h_m\left(\frac{x}{w_o}\right) \tag{7.2.16}$$

and the carrier density distribution as

$$N(x,z) = \sum_{m=0}^{M} b_m(z) h_m\left(\frac{x}{w_o}\right) \tag{7.2.17}$$

where $h_m\left(\frac{x}{w_o}\right) = C_m H_m\left(\frac{x}{w_o}\right)\exp\left(\frac{-x^2}{w_o^2}\right)$ are the HG basis functions, with $C_m = \frac{1}{\sqrt{\pi 2^m m! \sqrt{w}}}$ the normalising constant, w_o the (constant) width parameter; $a_m^{\pm}(z)$ and $b_m(z)$ are the expansion coefficients for the field and the carrier profile, respectively.

The transverse axis is discretised by making use of the collocation method [Chapter 4, Section 4; and Chapter 6, Section 2]. The same discretisation is used for both the field and the carrier equations, and hence the same w_o and the same number of expansion terms is considered in (7.2.16) and (7.2.17).

By substitution of (7.2.16) and (7.2.17) into equations (7.2.4), (7.2.5) and (7.2.14), respectively, two sets of coupled ordinary equations are derived: one is for the expansion coefficients of the field, $a_m^{\pm}(z)$, [Chapter 4] and the other for the expansion coefficients of the carrier density profile, $b_m(z)$, [Chapter 6].

The sets of coupled equations for the field are

$$\frac{d\underline{f}_{\pm}(x,z)}{dz} = \mp\frac{i}{2p}\left(\underline{\underline{S}}(z) - p^2\underline{\underline{1}}\right)\underline{f}_{\pm}(x,z) \tag{7.2.18}$$

where $m = 1, 2, ..., M$, $\underline{f}_{\pm} = \{f_{\pm}(x_i,z), i = 1, 2, ..., M\}$, $\underline{\underline{1}}$ the unity matrix and $\underline{\underline{S}}(z)$ as defined in Appendix 4.5. For the carrier profile

$$\underline{N} = -\underline{\underline{T}}^{-1}\underline{J} \tag{7.2.19}$$

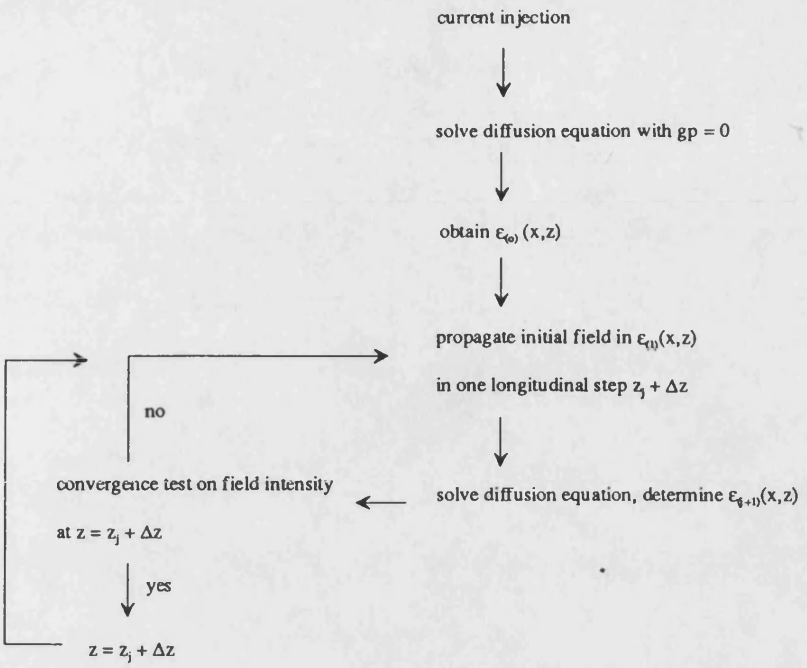
with $\underline{N} = \{N(x_i),\ i = 1, 2, \ldots, M\}$ and $\underline{\underline{T}}$ the invertible matrix as defined in Appendix 6.2.

2.5 Numerical details

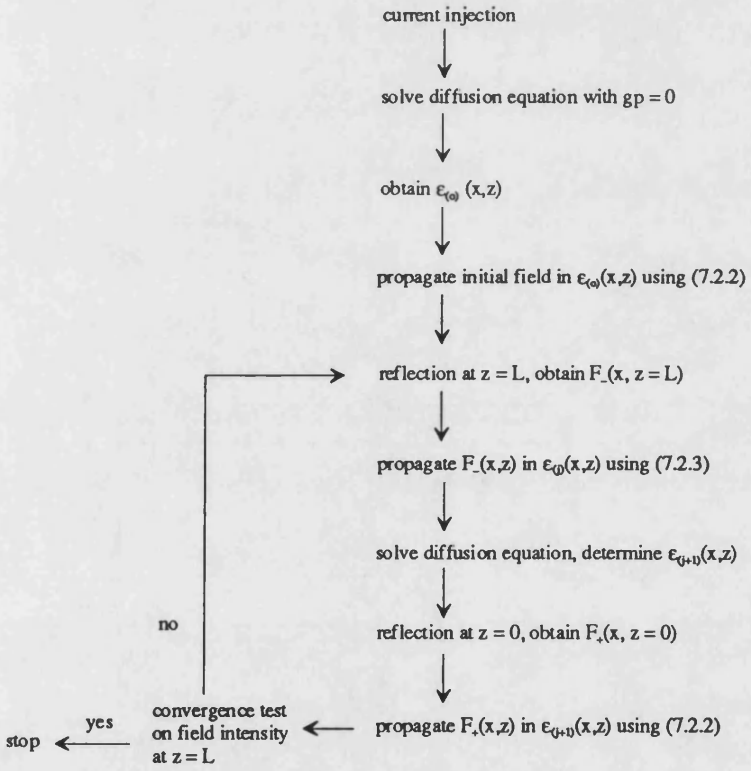
The convenient matrix form of the resulting sets of equations makes it possible to use standard techniques of solutions. In the present formulation the Runge-Kutta variable step - variable order Adam's method has been used to solve the set of first order differential equations for the field, (7.2.18).

Travelling-wave amplifiers and lasers are modelled using two different iterative schemes. In the former case the solution is locally (i.e., at each longitudinal step) self-consistent, Fig. 7.2.2 a), since the field travels along the structure just once.

In the laser model the field is propagated between the cavity mirrors until a stationary field configuration is achieved, which represents the cavity output mode (in the steady state analysis). The (complex) refractive index distribution in the medium changes as the field evolves. At each iteration the solution (optical intensity profile) is tested at the facets for convergence. The iteration scheme used in the present formulation for the laser, which is also known as Fox-Li method, [16], is given in the flow chart of Fig. 7.2.2 b). Note that in the diffusion equation (7.2.14) use is made of $S_{av}(x,z) \propto \left(|f_+(x,z)|^2 + |f_-(x,z)|^2\right)$ which corresponds to the summation of the powers of the forward travelling and the reverse travelling wave. With reference to Fig. 7.2.2 b), during each single pass the power of the travelling field is stored so that in the next (single) pass S_{av} (x,z) can be calculated by retrieving the power carried by the field travelling in the opposite direction.



a) travelling-wave amplifier



b) laser

Fig. 7.2.2: Schematic for the iteration procedure used with the self-consistent HGCM model for a) travelling-wave amplifiers, b) laser devices; (gp indicates the stimulated emission term in the carrier diffusion equation, (7.2.14); $\varepsilon_{(j)}(x, z)$ the complex dielectric distribution calculated in the j-th iteration; Δz the longitudinal step).

3. Semiconductor amplifier and lasers: HGCM computed results

Consider in the present section a linearly tapered geometry semiconductor structure as shown in Fig. 7.2.1 c), whose dimensions are specified in Tab. 7.3.1. In the following subsections the same structure will be used to model a travelling-wave flared amplifier, Sections 3.1, and a tapered laser, Section 3.2.

For the results presented in the following subsections the same parameters have been used with the HGCM, i.e., M = 99, w_o = 1.3μm, the step in the longitudinal direction Δz = 5μm, unless otherwise specified.

Tab. 7.3.1: Parameters defining the tapered geometry structure of Fig. 7.2.1 c)

Parameter	value
$2a_o$	3μm
$2a_{fin}$	20μm
L	500μm
n_1	3.33
n_2	3.325
λ_o	0.86μm

3.1 Tapered geometry travelling-wave amplifier

The reflectivity at both facets of the travelling-wave amplifier is assumed to be zero, the other parameters are summarised in Tab. 7.3.2. Assuming the input optical field to be the fundamental mode of the local slab at z = 0, the solution is calculated with the HGCM using the self-consistent scheme of Fig. 7.2.2 a).

3.1.1 Comparison with the mode matching technique

The results obtained with the HGCM for the travelling-wave amplifier may be compared with those calculated with another method based on the local mode expansion technique, described in [17] (referred to as DMM). Tab 7.3.1 gives the parameters for the devices used for the comparison.

The near fields calculated with the two methods are plotted in Fig. 7.3.1. The agreement between the two computed field profiles is quite satisfactory.

Whereas in the passive tapered structure the power carried by the field is conserved during propagation, in the travelling-wave amplifier it increases due to the gain introduced by the injected carriers, as shown in Fig. 7.3.2, where the gain rates obtained with the two methods are compared.

Note that the output power calculated with the HGCM is larger than that obtained with the DMM, which may be due to i) the fact that the discretisation procedures are different in the two methods, i.e., the structure simulated with the HGCM does not correspond to the one simulated with the DMM, ii) the approximation of neglecting the contributions of the radiation (local) modes in the DMM.

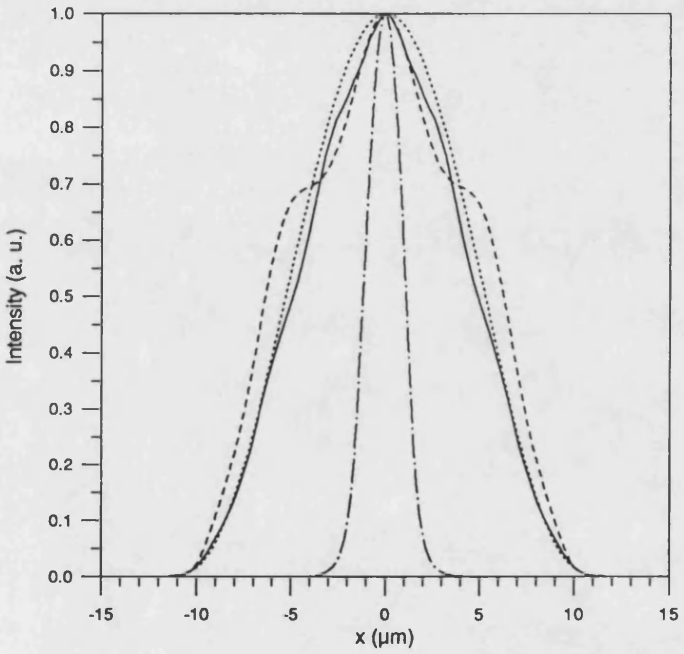


Fig. 7.3.1: Linearly tapered travelling-wave amplifier, near field (z = L): solid line: HGCM; broken line: DMM; dotted-broken line: input field; dotted line: fundamental mode at z = L (J_o = 1.5 10^{-5} A/μm², P_i = 5mW).

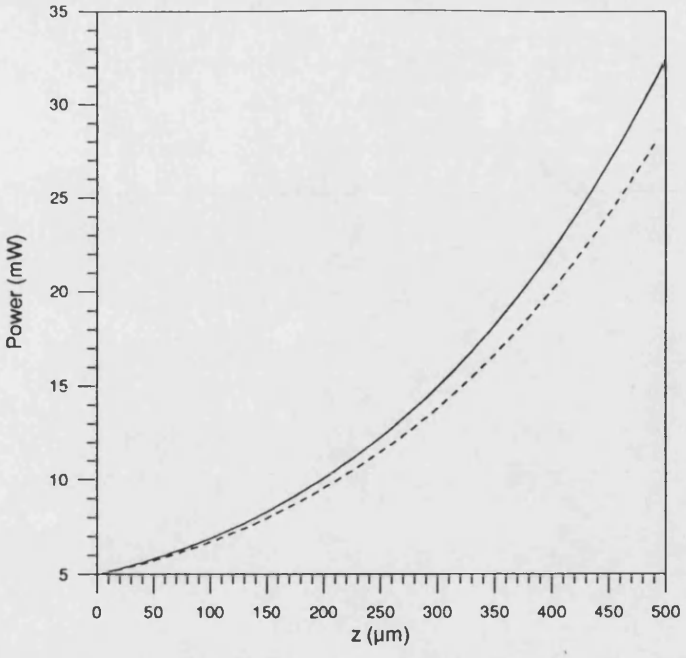


Fig. 7.3.2: Tapered amplifier: single pass gain for J_o = 1.5 10^{-5} A/μm², and input power P_i = 5mW: solid line: HGCM (P_{out} = 32mW); broken line: DMM (P_{out} = 28mW).

Tab. 7.3.2: Parameters used for modelling the tapered geometry amplifier

parameter	value	parameter	value
R_1	0	α_{cl}	0.25 10^{-3} μm^{-1}
R_2	0	S_{sat}	0.2 10^{3} Wμm^{-2}
(+) Γ	0.07307	D	3.5 10^{9} $\mu m^2 s^{-1}$
a_c	1.4 10^{-9} μm^3	B_r	1.4 10^{2} $\mu m^3 s^{-1}$
α	1.5 10^{-8} μm^2	γ	0.0 $\mu m^6 s^{-1}$
N_{tr}	10^{6} μm^{-3}	n_o	1.0 10^{4} μm^{-3}
α_{sc}	0.0 μm^{-1}	d	0.02 μm
α_a	1.0 10^{-3} μm^{-1}	q	1.6 10^{-19} C

(+) The vertical confinement factor used in the calculations corresponds to typical Quantum Well active layer

In Fig. 7.3.3 the light-current characteristics for the amplifier obtained with the two methods are shown. As also observed with other methods of solution, [17], for high power densities it has been found that the HGCM iteration scheme may not reach convergence. It is observed that the output beam characteristics do not change with varying J_o or P_{in}, which can be explained by the fact that the effect of the field-carrier

interaction on the refractive index is small in amplifiers made of Quantum Well material.

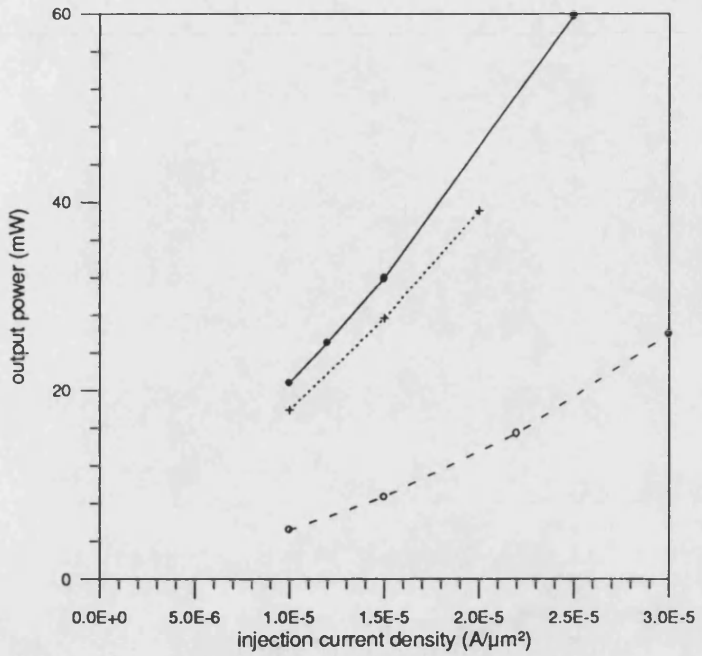


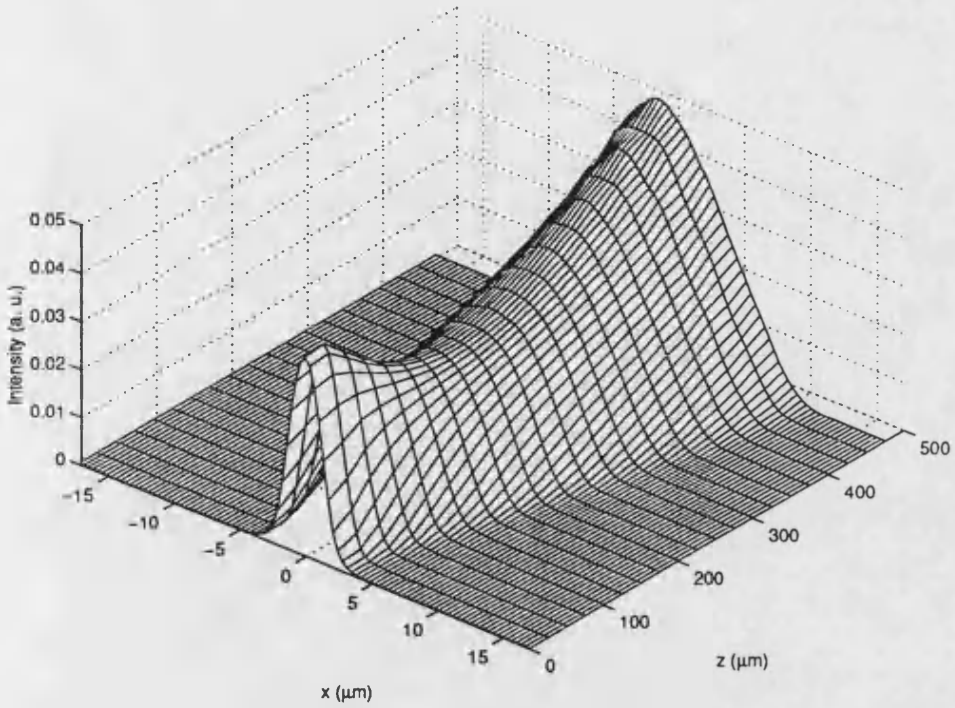
Fig. 7.3.3: Light-current characteristics for the tapered geometry travelling wave-amplifier: broken line: P_{in} = 1mW; solid line: P_{in} = 5mW; dotted line: DMM with P_{in} = 5mW.

3.1.2 Forward and reverse travelling fields

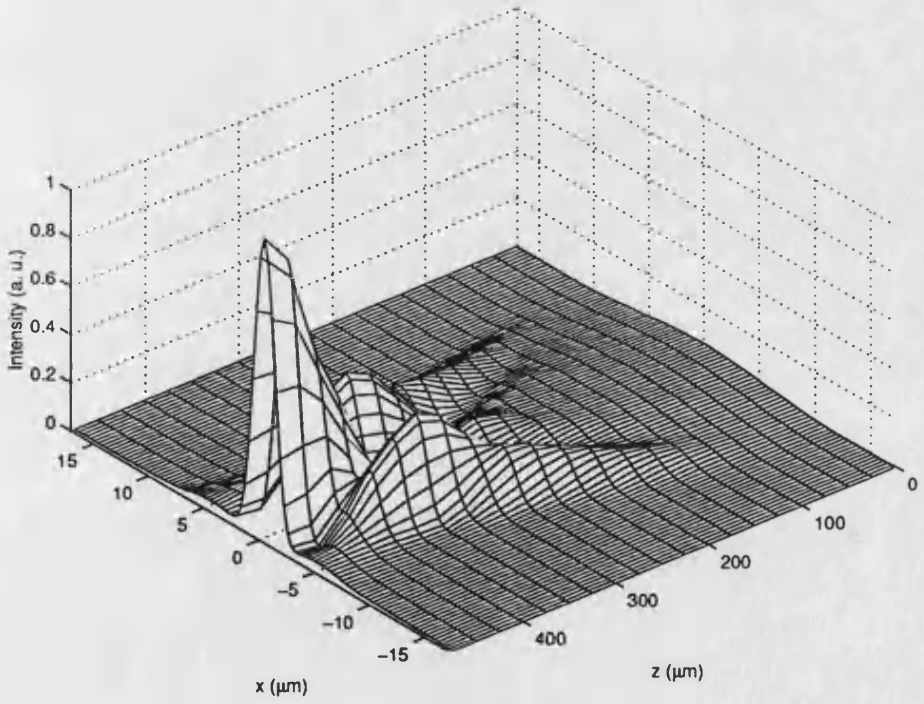
The last set of results for the travelling-wave amplifier shows the difference between the forward and reverse travelling fields. This has been done in order to attain information on the iteration process that will be used in the laser model. The field shown in Fig. 7.3.4 a) has been obtained by solving equation (7.2.2), while that in Fig. 7.3.4 b) by solving equation (7.2.3). In Fig. 7.3.4 a) the solution of the diffusion equation shows the interesting feature that the inversion population density in the narrow region is below transparency, for the value of injection current used in the calculations, so that the optical field intensity initially reduces as it propagates along z because of absorption. Beyond the longitudinal position where the minimum intensity occurs the effect of diffusion is reduced so that the carrier density is now above transparency and hence the field intensity sees optical gain.

Note also that the initial field in Fig. 7.3.4 b) is the output field which has been obtained after forward propagation, but the scaling changes from Fig. 7.3.4 a) to Fig.

7.3.4 b). The field obtained in Fig. 7.3.4 a) and b) are plotted for z = 0 in Fig. 7.3.5 to highlight the differences between the two profiles. The radiation appearing at the narrow end of the taper for backward propagation is expected to have a significant effect on the modelling of laser devices, as discussed in Section 3.2 of this Chapter.



a) forward propagation



b) backward propagation

Fig. 7.3.4: Field propagation in the tapered travelling-wave amplifier: a) forward, b) reverse travelling field (J_o = 1.0 10^{-5} A/μm², P_{in} = 0.1mW).

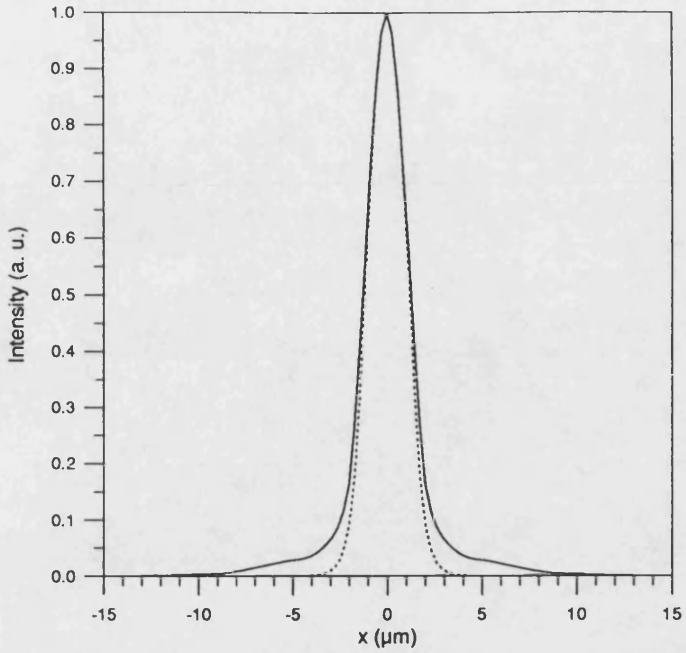


Fig. 7.3.5: Field propagation in the tapered travelling-wave amplifier: comparison of the fields obtained in Fig. 7.3.4 a) and b) at z= 0.

3.2 Linearly tapered laser

The device structure to be modelled in this subsection is shown in Fig. 7.2.1 c). The tapered laser modelled in this section is characterised by the same parameters used for the travelling-wave amplifier, and hence refer to Tab. 7.3.1 and Tab. 7.3.2 for the device parameters. For computing laser operation the facet reflectivities have been assumed to be equal and taken to be $R_1 = R_2 = 0.3$.

In the various test cases discussed in the following paragraphs, unless otherwise stated, the injection current density is $J_o = 1.5\ 10^{-5} A/\mu m^2$, which corresponds to $1.5 \times J_{th}$, with J_{th} the threshold current density of the device of interest.

3.2.1 $a_c = 0$, $\alpha = 0$

The first set of results refers to a structure where the effect of the injected carries on both the gain and the refractive index has been suppressed, i.e., with reference to equations (7.2.8) and (7.2.12), $a_c = 0$, $\alpha = 0$. Also the scattering losses and the losses due to the free carriers, equation (7.2.7), are neglected in the present section and the reflectivity of the facets are assumed to be unity. This effectively corresponds to a passive device since both the gain and the losses are neglected.

The near field changes at each iteration; as an example, the near field obtained after 15 round-trip iterations is shown in Fig. 7.3.6, from which it can be noted that most of the field leaks out of the rib.

The quantity $\int_{x_1}^{x_M} |F(x,L)|^2 dx \propto$ Power carried by the field at z = L, where (x_1, x_M) is the computational window used with the HGCM, remains constant at each iteration since all the losses and the gain factor have been set to zero, Fig. 7.3.7.

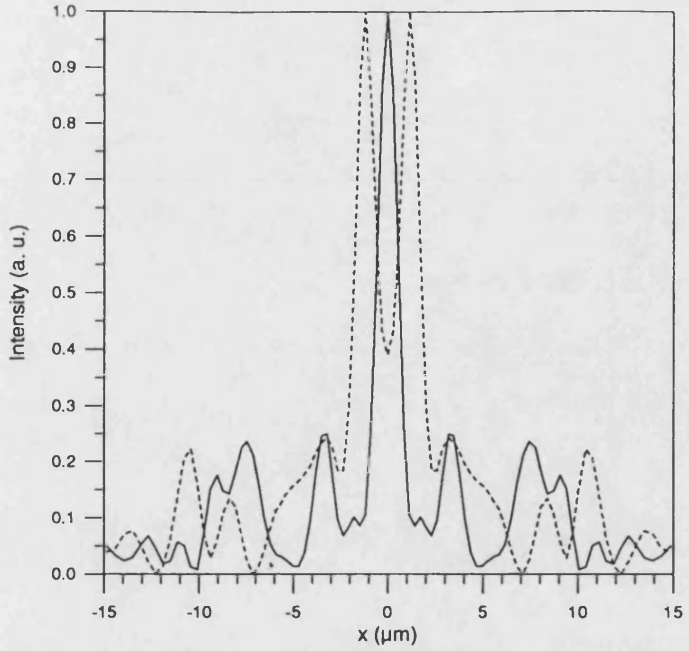


Fig. 7.3.6: Near field for tapered laser at z = 0 (broken line), and at z = L (continuous line).

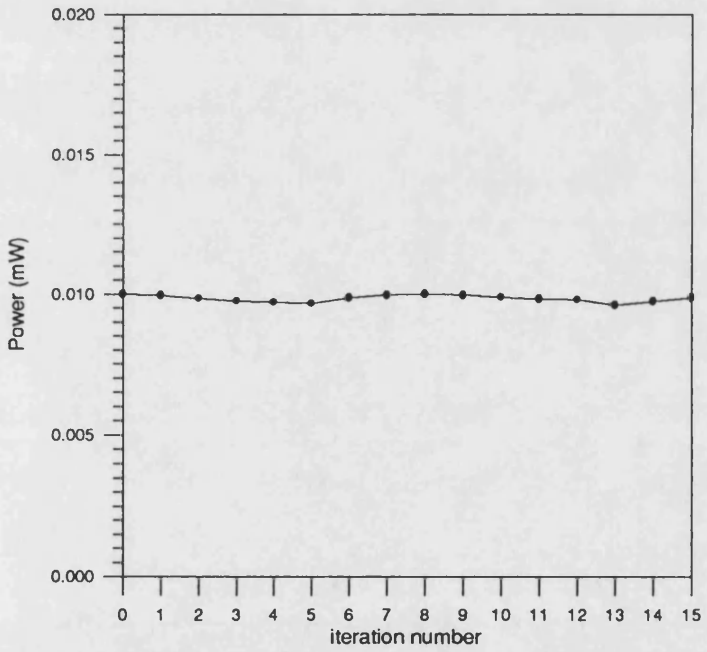


Fig. 7.3.7: Output power after each iteration.

3.2.2 $a_c = 0$, $\alpha \neq 0$

The following results have been calculated for the case $a_c = 0$, but $\alpha \neq 0$ ($\alpha = 1.5\ 10^{-8}$ μm^2), in equations (7.2.8) and (7.2.12), respectively. This corresponds to the situation in which the refractive index is not affected by the presence of the injected carriers, but the carrier distribution is modified by the optical field in the device, as shown in Fig. 7.3.8.

At each iteration the near field profile, Fig. 7.3.9, does not change, apart from a scaling factor due to gain, which is an indication of stability in the field solution. It is in fact observed that the same solution is obtained even after starting with a different initial field in the iterative process. However, from Fig. 7.3.10 it can be seen that the output power seems to be reaching a stationary value after quite a large number of round trip iterations.

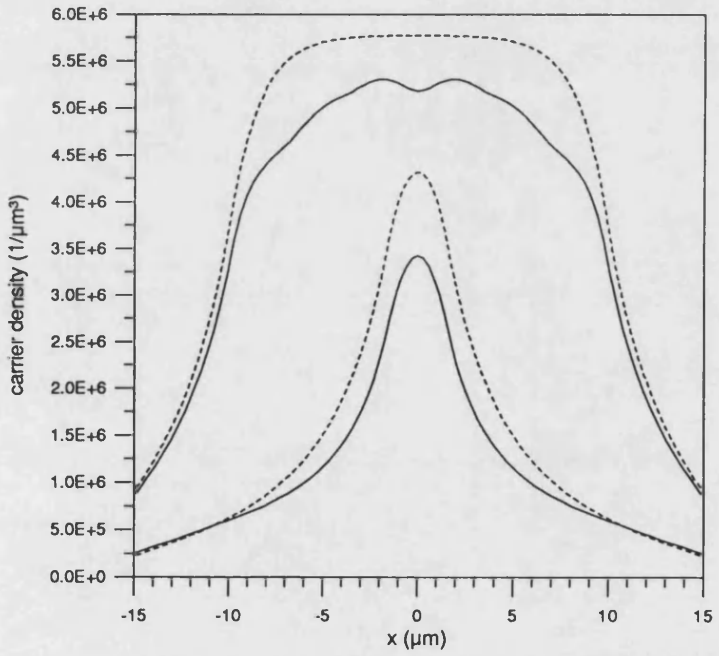


Fig. 7.3.8: Carrier density distribution at the narrow and wide ends of the linearly tapered laser: broken line: in the absence of the optical field in the cavity; continuous line: solution after reaching self-consistency.

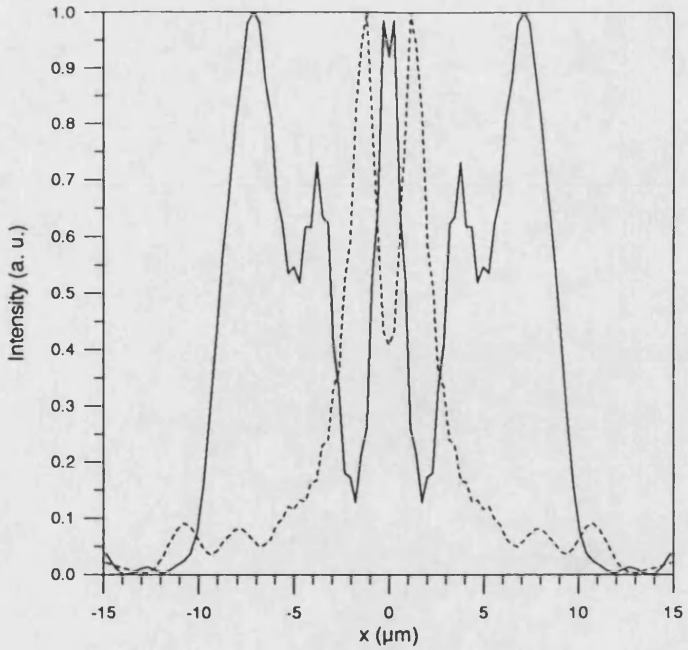


Fig. 7.3.9: Near field for the tapered laser at z = 0 (broken line) and at z = L (continuous line).

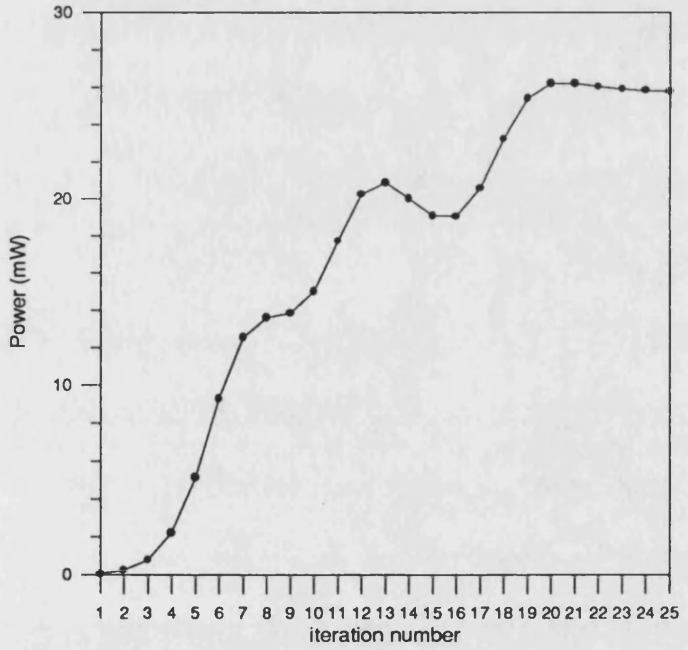
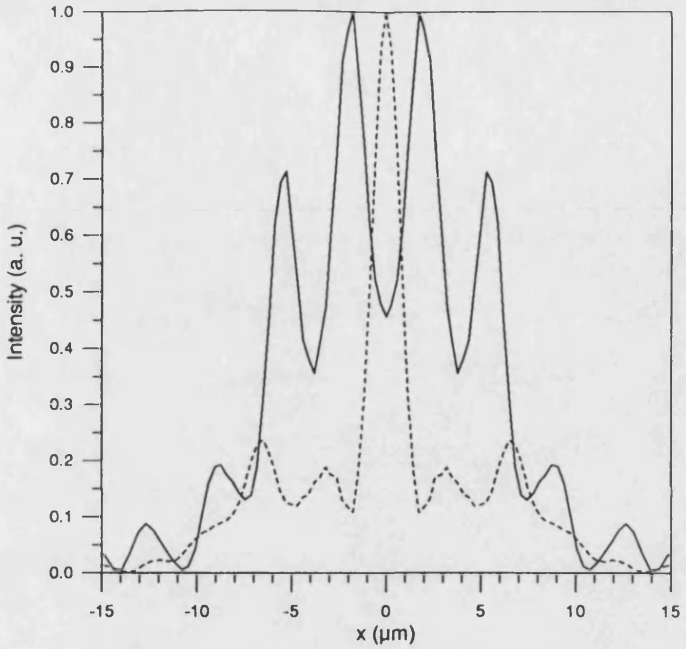


Fig. 7.3.10: Output power obtained after each round-trip iteration.

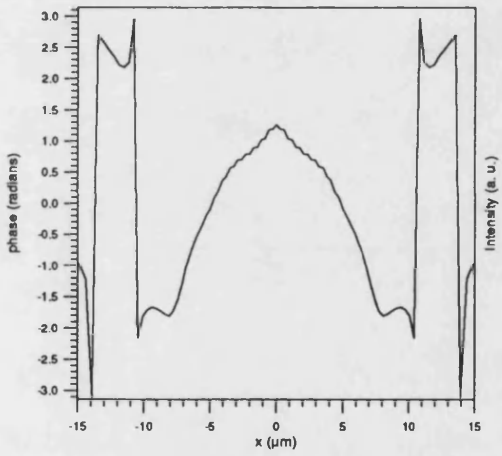
3.2.3 $a_c \neq 0$, $\alpha \neq 0$

In this subsection the self-consistent solution is calculated for the case in which both a_c and α, in equations (7.2.8) and (7.2.12), are finite, i.e., $\alpha = 1.5\ 10^{-8}$ μm^2 and $a_c = 1.4\ 10^{-9}$ μm^3. Both the near field, Fig. 7.3.11, and the carrier density distribution, Fig. 7.3.12 a), are different from those presented in the previous cases, showing the effect of the change in refractive index due to the carriers ($a_c > 0$), Fig. 7.3.12 b).

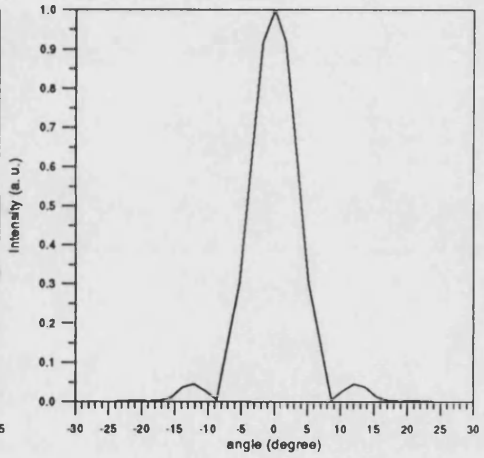
The output power obtained after each round trip iteration is plotted in Fig. 7.3.13 for different injection currents. From Fig. 7.3.13 it can be seen that at relatively high injection currents an oscillatory pattern appears in the output power (in Fig. 7.3.13 the scale has been explicitly chosen so that it will be possible to compare this result with the corresponding one obtained for a parabolic laser, which is discussed in Section 4). This oscillatory pattern in the power may be dependent on i) the accuracy achievable with the matrix formulation of the iterative scheme, or ii) the geometry of the structure. This point needs further investigation.



a) near field

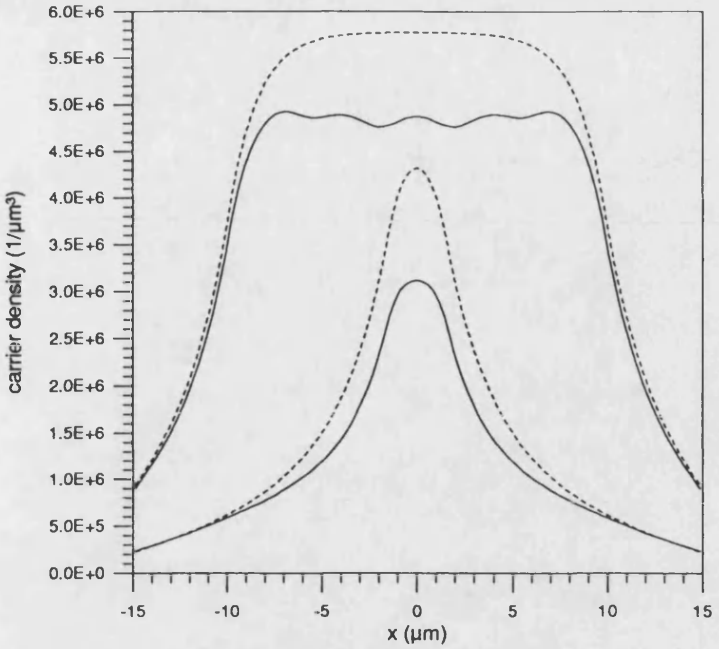


b) phase front

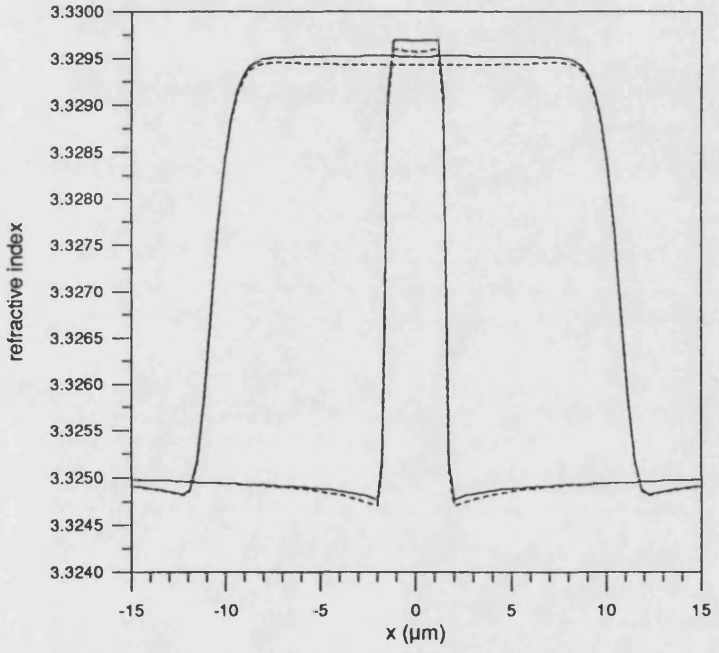


c) far field intensity profile

Fig. 7.3.11: Tapered laser: a) near field at z = 0 (broken line) and at z = L (continuous line); b) phase front at z = L; c) far field intensity profile.



a) carrier distribution at z = 0 and at z = L



b) refractive index profile at z = 0 and at z = L

Fig. 7.3.12: Tapered laser: broken line: without the optical field in the cavity; continuous line: self-consistent solution.

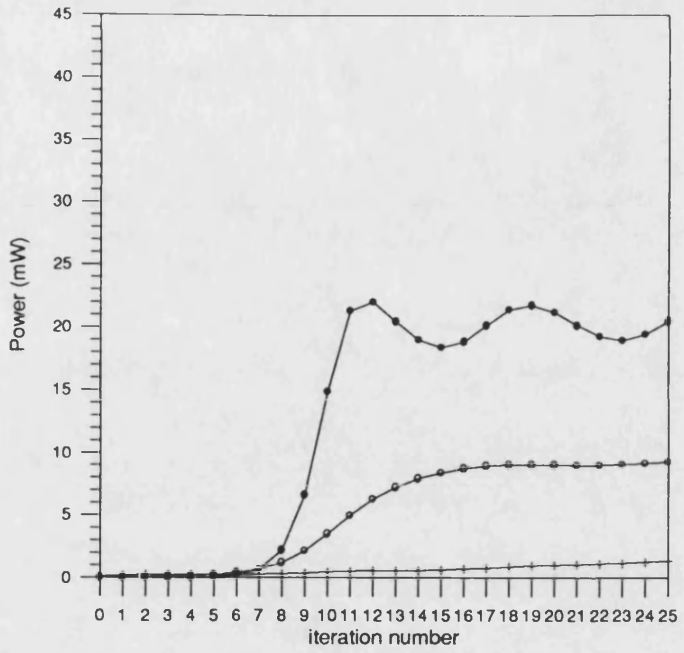


Fig. 7.3.13: Tapered laser: output power obtained after each round-trip iteration: (+) $J_o = 1.0\ 10^{-5}$ A/μm², (∘) $J_o = 1.2\ 10^{-5}$ A/μm², (•) $J_o = 1.5\ 10^{-5}$ A/μm².

3.2.4 $a_c \neq 0$, $\alpha \neq 0$, M = 51, w_o = 1.8μm

The results presented in subsection 3.2.3 have been repeated using M = 51 and width parameter w_o = 1.8μm in the HGCM calculations. It is found that the results closely reproduce those obtained with M = 99 and w_o = 1.3μm, the oscillations in the output power are also reproduced.

Comparing Fig. 7.3.14, where the near field obtained with M = 51 is shown, with Fig. 7.3.11 a) it is seen that the HGCM with M = 51 produces a self-consistent solution which is in qualitatively good agreement with the one obtained with M = 99, however, a larger number of collocation points provides a more detailed near field, which is desirable.

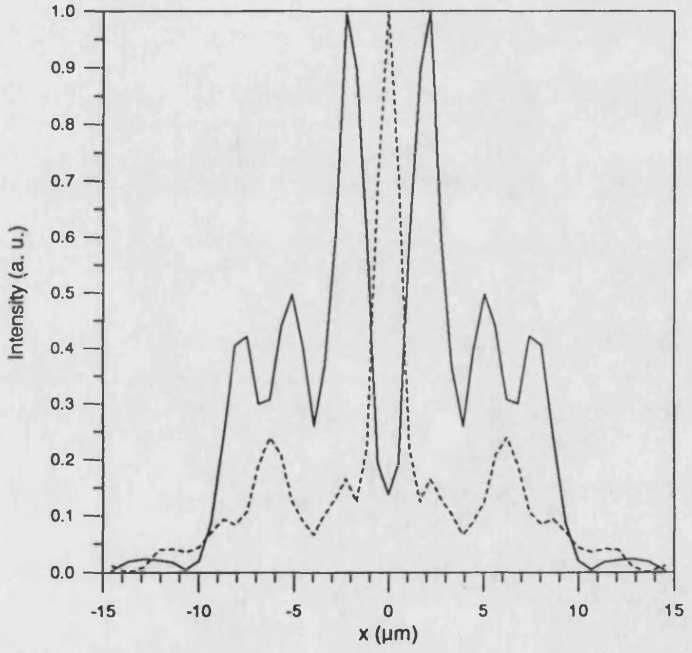


Fig. 7.3.14: Near field for the tapered laser: broken line: near field at z = 0; continuous line: near field at z = L.

3.2.5 $\alpha \neq 0$, $a_c \neq 0$, M = 99, w_o = 1.3μm, with deflectors

The last set of results of this section refer to the tapered laser discussed in the subsection 3.2.3 with an additional modification. It has been noted experimentally that the inclusion of deflectors at the narrow section of tapered lasers, [18], greatly improves the quality of the output beam characteristics. The deflectors in fact act as a spatial filter, preventing the field that leaks outside the rib (radiation) to be reflected and coupled back into the taper.

With reference to Fig. 7.3.15, preliminary results have been obtained representing the deflectors in the model modifying the facet reflectivity at z = 0 from $R = R_1 \; \forall x$ to

$$R = \begin{cases} R_1 & |x| \leq 2a_o \\ 0 & |x| \geq 2a_o \end{cases} \tag{7.3.1}$$

Equation (7.3.1) has been thus replaced by the following (analytic) supergaussian distribution to avoid numerical inaccuracies due to the collocation method,

$$R(x) = R_1 \exp\left\{-\frac{1}{2}\left(\frac{x}{(2a_o)}\right)^q\right\} \tag{7.3.2}$$

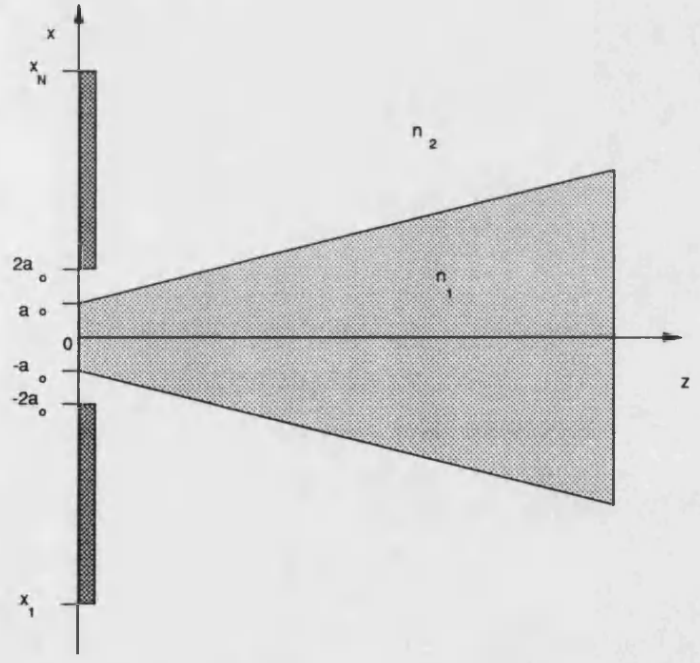


Fig: 7.3.15: Model for the inclusion of deflectors (darker shaded region at z = 0) in the tapered laser; (x_1, x_M) transverse range specified by the collocation points; $2a_o$ is the narrower width of the taper.

With this device model the power output converges with the iteration number, Fig. 7.3.16, thereby demonstrating that the presence of the deflectors improves the results, but it is found that the near field is significantly altered, Fig. 7.3.17. The 'noise' appearing in the near field profile of Fig. 7.3.17 a) is, however, to be expected because of the truncation of the field due to the deflectors, equation (7.3.1); as demonstrated by the use of the supergaussian representation of the facet reflectivity, equation (7.3.2), Fig. 7.3.17 b).

Finally, note that the carrier distribution, Fig. 7.3.18, has changed as compared to that shown in Fig. 7.3.12 a). By comparing Fig. 7.3.18 with Fig. 7.3.12 a) it is possible to notice that the characteristic 'hole-burning' effect on the carriers has changed as a consequence of the change in the field intensity profile.

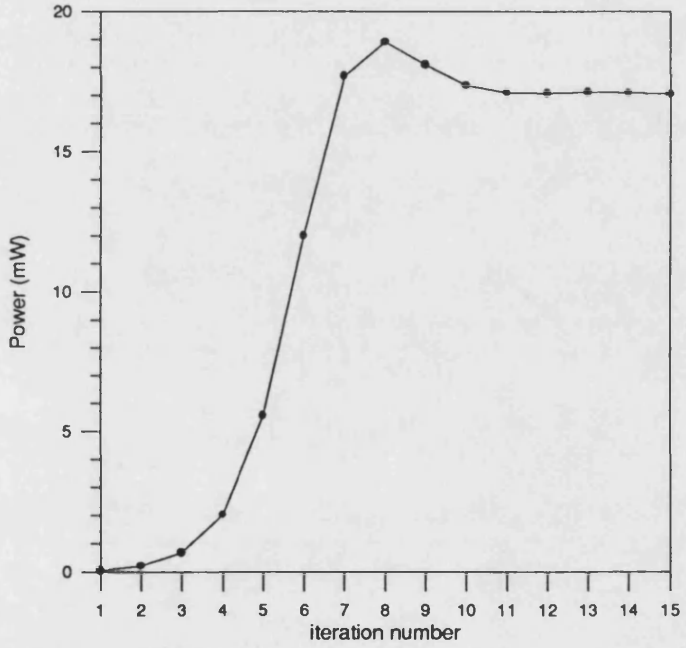


Fig. 7.3.16: Power as a function of the iteration number for the tapered laser with deflectors.

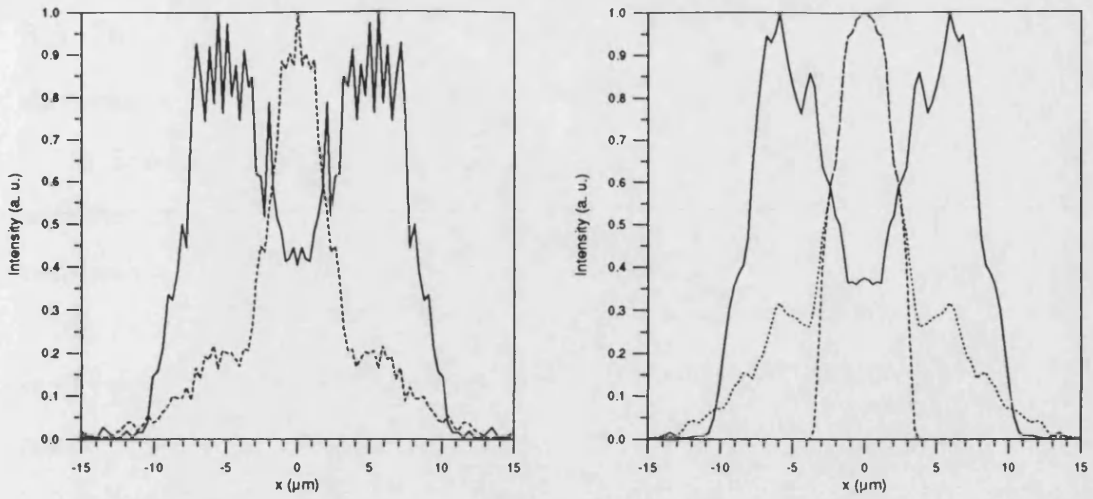


Fig. 7.3.17: Near field profile in tapered laser with deflectors using the representation of a) equation (7.3.1), and b) equation (7.3.2): broken line: at z = 0; continuous line: at z = L; in b) dotted line: field after reflection at z = 0.

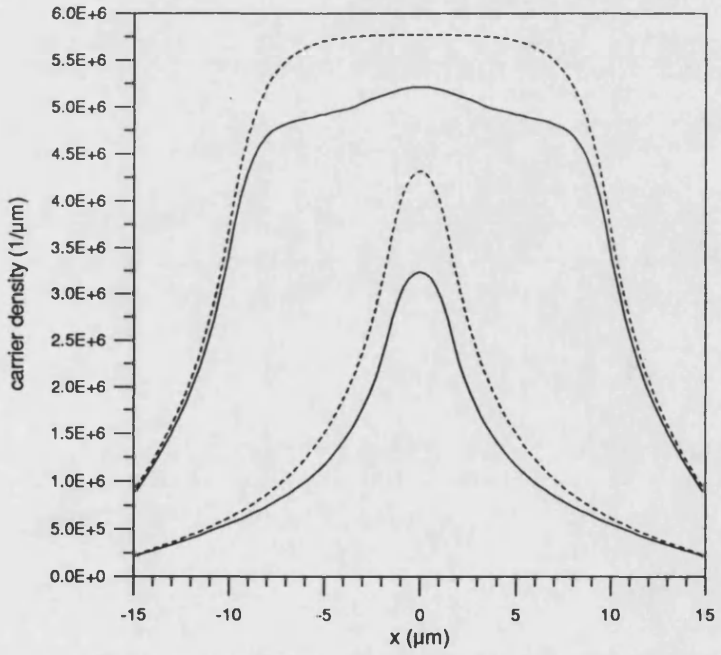


Fig. 7.3.18: Carrier distribution at the narrow and wide ends of the tapered laser: broken line: without the optical field in the device; continuous line: solution after reaching self-consistency.

3.3 Discussion of the HGCM results for linearly tapered active devices

In Section 3.1 results calculated with the HGCM for a flared travelling-wave amplifier are presented and compared with those produced with another self-consistent model, the DMM, showing satisfactory agreement.

In the series of results presented in Sections 3.2.1 to 3.2.3, the parameters a_c and α of equations (7.2.8) and (7.2.12), have been introduced one at a time to observe various stages of self-consistency of the HGCM scheme.

In Section 3.2.4 a smaller (M = 51) number of collocation points is used to model the linear laser obtaining good qualitative agreement with the results calculated with M = 99, Section 3.2.3. However, a larger number of collocation points is desirable to produce detailed field characteristics.

The characteristic feature of the HGCM scheme of including the contribution of the radiation (local) mode spectrum is found to be of fundamental importance in the modelling of semiconductor active optical devices. The results shown in Fig. 7.3.5 c) demonstrate that at the narrow end of the taper, Fig. 7.2.1 c), the field partially radiates outside the rib. It is thus argued that in the iterative process to find the self-consistent solution for lasers, the reflection of the 'radiation tails', which thus couples back into the tapered device, significantly affects the self-consistent field profile, as shown in Fig. 7.3.11 a).

The elimination of the 'radiation tails' by the incorporation of a spatial filter in the model (Section 3.2.5), has the effect of changing the self-consistent field profile and, consequently, also the carrier density profile.

It is to be noted that the HGCM near threshold converges to a self-consistent solution, however at high injection current densities it may fail to reach convergence because with large power densities the effects of stimulated recombination are dominant and may destabilise the iterative scheme. This feature as also been noted in other numerical iterative schemes, [17].

No comparisons are shown in this section for linearly tapered lasers since it has not been possible to readily develop an alternative model for comparison. Note in this respect that the DMM, which has been used as a comparison for the travelling-wave amplifier, is not amenable for the self-consistent solution of linearly tapered lasers

since the backward propagating higher order modes cut off as they approach the narrow end of the taper, and hence the description of the field is not sufficiently accurate.

4. Parabolically tapered laser

The need for achieving high power semiconductor lasers but at the same time with good far field characteristics for applications in medicine, communications, laser printing, etc., has lead to the development of parabolically tapered devices. It has been shown that parabolic taper lasers produce very desirable operational characteristics yielding moderately large output power, reduced optical power density at the broad facet and narrow far-fields (without a correcting lens) for output facet widths up to about 30 μm, [19]. In general the theoretical comparison between linear and parabolic devices indicates that in the latter structures the near field is wider, hence making a better use of the active semiconductor material; the far field is typically narrower and the low power density distributed inside the device reduces the potential for catastrophic optical damage of the facets. Similar features are also observed experimentally, [19].

In this respect the results calculated with the HGCM for the parabolic tapered laser can be compared with those discussed in the previous section for linearly tapered structures.

The first observation, by comparing Fig. 7.4.1 a) with Fig. 7.3.11 a) is that the field leaking out of the rib is significantly reduced in the parabolic structure. This is expected since it has been shown [20] that the coupling to the fundamental local mode is maximised by the parabolic rib profile. Further evidence that the reverse travelling field is contained almost entirely within the parabolic structure is given by observing that the inclusion of deflectors, as in the linear taper, Fig. 7.3.15, does not significantly change the overall field distribution in this case - unlike in the linear taper, [Section 3.2.5, Fig. 7.3.17, in this chapter]. The phase front, Fig. 7.4.1 b) and Fig. 7.3.11 b), also appears to be 'flatter', which, consequently produces a narrower far field profile, Fig. 7.4.1 c) and Fig. 7.3.11 c).

Also note that the power output is considerably higher than that achieved with the linear structure, as shown in Fig. 7.4.2 for three values of the injection current density.

Further, by comparing Fig. 7.4.2 with Fig. 7.3.13, where the same scale has been used, the oscillation with iteration number in the output power that was observed in the modelling of the linear taper is now no longer observed.

To complete the picture the carrier density and the refractive index distributions produced in the parabolic device are shown in Fig. 7.4.3.

Because the DMM can be used to model the parabolic tapered laser (since the backward travelling fields are reasonably accountable in this case), a comparison of the results obtained with the DMM and with the HGCM is possible. The results calculated with the DMM using the same parameters are also presented in Fig. 7.4.1 for the near field profile, the phase front at the wide facet and for the far field. The very good agreement between the two sets of results is most satisfying, considering that the two methods of analysis are quite different.

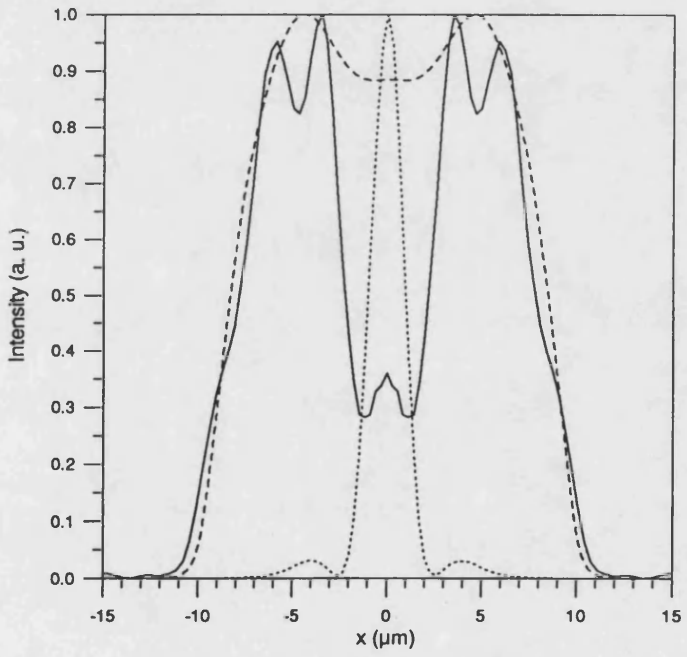
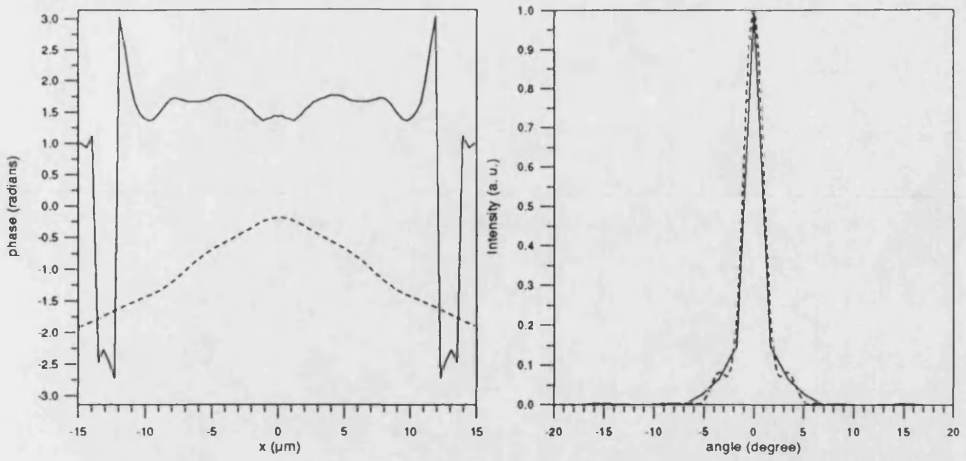


Fig. 7.4.1 (continued) : a) near field



b) phase front c) far field intensity profile

Fig. 7.4.1 (continued): Parabolically tapered laser: a) near field: dotted line: at z = 0; solid line: at z = L; b) phase front at z = L; c) far field intensity profile. In all three graphs: solid line: HGCM; broken line: DMM.

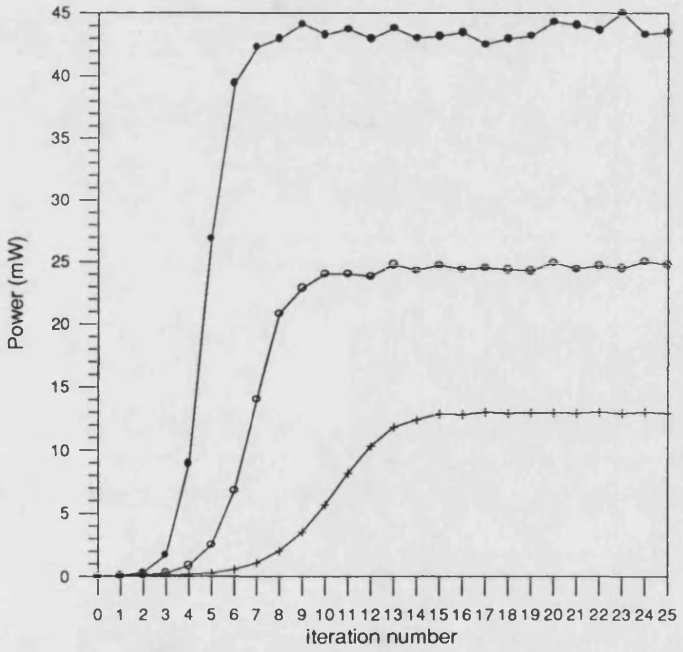
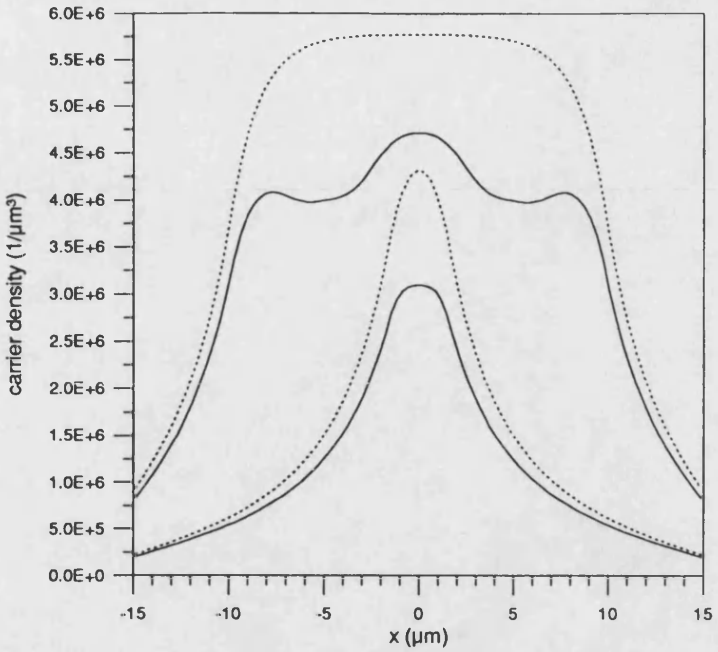
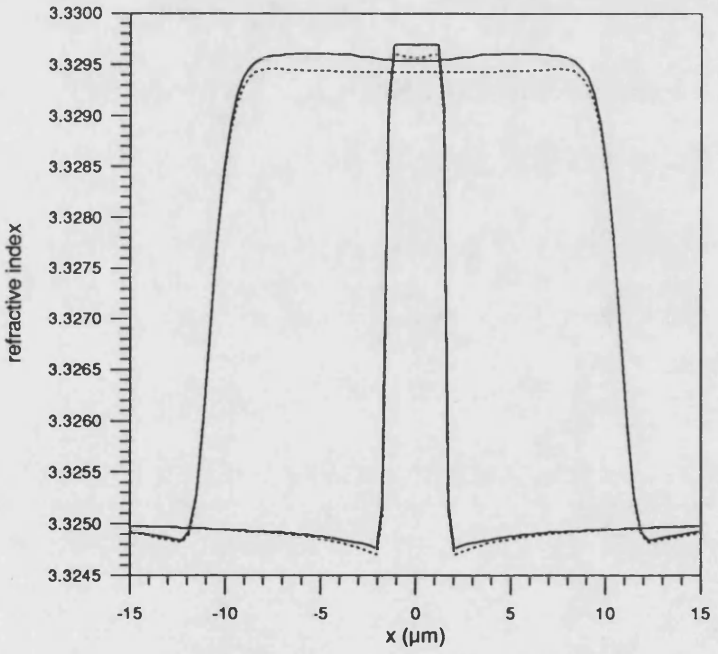


Fig. 7.4.2: Parabolically tapered laser: output power obtained after each round trip iteration: (+) $J_o = 1.0\ 10^{-5}$ A/μm², (◦) $J_o = 1.2\ 10^{-5}$ A/μm², (•) $J_o = 1.5\ 10^{-5}$ A/μm².



a) carrier distribution at z = 0 and at z = L



b) refractive index profile at z = 0 and at z = L

Fig. 7.4.3: Parabolically tapered laser: broken line: without the optical field in the cavity; continuous line: solution after reaching self-consistency.

5. Stripe laser

The objective of this section is to establish that the HGCM model can be used equally to analyse gain-guided devices as it has been to model index-guided devices. For this purpose a narrow stripe laser (with device parameters as in Tab. 7.5.1) is modelled, Fig. 7.5.1.

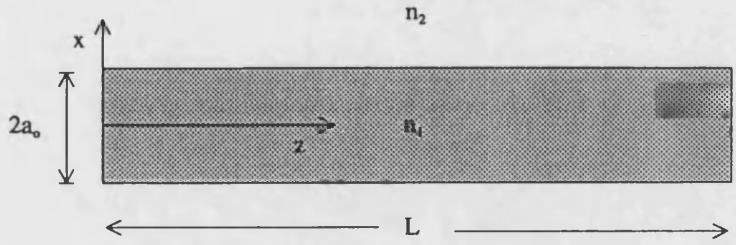


Fig. 7.5.1: Stripe laser geometry.

The device parameters listed in Tab. 7.5.1 are the same as in [21], [22] so that results from the HGCM may be compared with those presented in the above references. For the HGCM, consider M = 51 and w_o = 1.8μm.

Tab. 7.5.1: Parameters used for modelling stripe lasers

parameter	value	parameter	value
$2a_o$	5μm	α_{sc}	0.0 μm^{-1}
L	250μm	α_a	1.0 $10^{-3} \mu m^{-1}$
n_1	3.33	α_{cl}	0.25 $10^{-3} \mu m^{-1}$
n_2	3.33 – Δn	S_{sat}	0.2 10^3 Wμm^{-2}
Γ	0.2517	D	1.0 10^9 $\mu m^2/s$
a_c	1.4 10^{-9} μm	B_r	1.0 10^2 $\mu m^3/s$
α	1.3 10^{-8} μm^2	γ	0.0 μm^6 /s
N_{tr}	10^6 μm^{-3}	n_o	1.0 10^4 μm^{-3}
d	0.2 μm		

To model the index-guided structure consider the shaded region in Fig. 7.5.1 as characterised by a different refractive index ($n_1 > n_2$) such that the lateral index step is $\Delta n = n_1 - n_2 = 5.0\ 10^{-3}$ (which is sufficient to provide index-guiding). On the other

hand, for the gain-guided, stripe laser assume the shaded region to be the metal contact in a homogeneous ($n_1 = n_2$) medium, and hence $\Delta n = 0$.

The near fields obtained after simulating the index- and gain-guided situations with the HGCM are compared in Fig. 7.5.2. As discussed in [21], [22] the field is wider in the gain-guided structure, while in the index-guided stripe laser the near field is expected to be narrower.

As expected, [14], the current threshold increases for the gain-guided device, Fig. 7.5.3 a), b). The refractive index distribution and the carrier density profile before and after achieving self-consistent solutions are shown in Fig. 7.5.4 a), b), c) and d) for the index- and the gain-guided devices, respectively. From Fig. 7.5.3 it is seen that $\Delta n = 5.0\ 10^{-3}$ is sufficiently large to ensure index-guiding, thereby preventing the effect of the injected carriers to bleach the lateral index step.

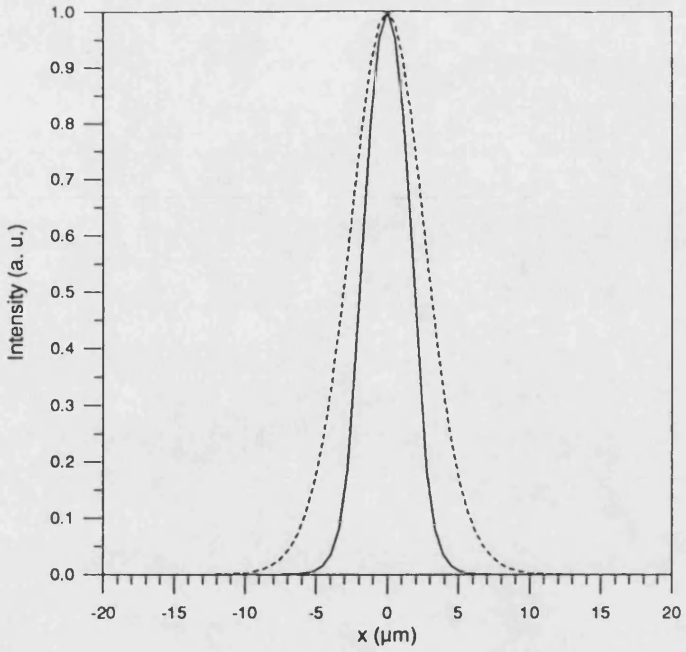
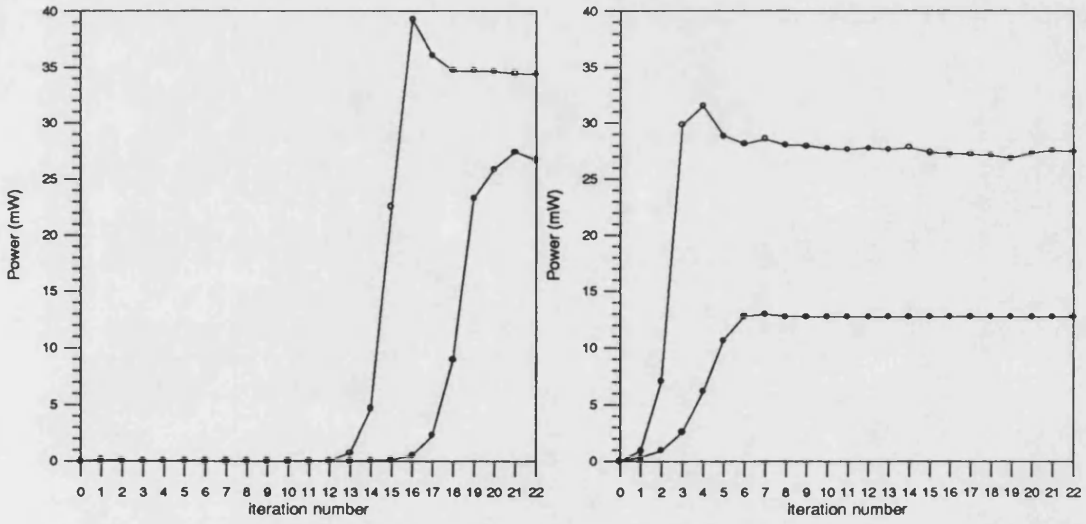


Fig. 7.5.2.: Near field: comparison between the gain-guided (broken line) and the index-guided stripe laser (continuous line).



a) gain-guided device **b) index-guided laser**

Fig. 7.5.3: Power calculated at each iteration for different injection current densities:

a) gain-guided device: (o) $J_o = 8.0\ 10^{-5}$ A/μm², (•) $J_o = 9.0\ 10^{-5}$ A/μm²;

b) index-guided device: (o) $J_o = 5.0\ 10^{-5}$ A/μm², (•) $J_o = 8.0\ 10^{-5}$ A/μm².

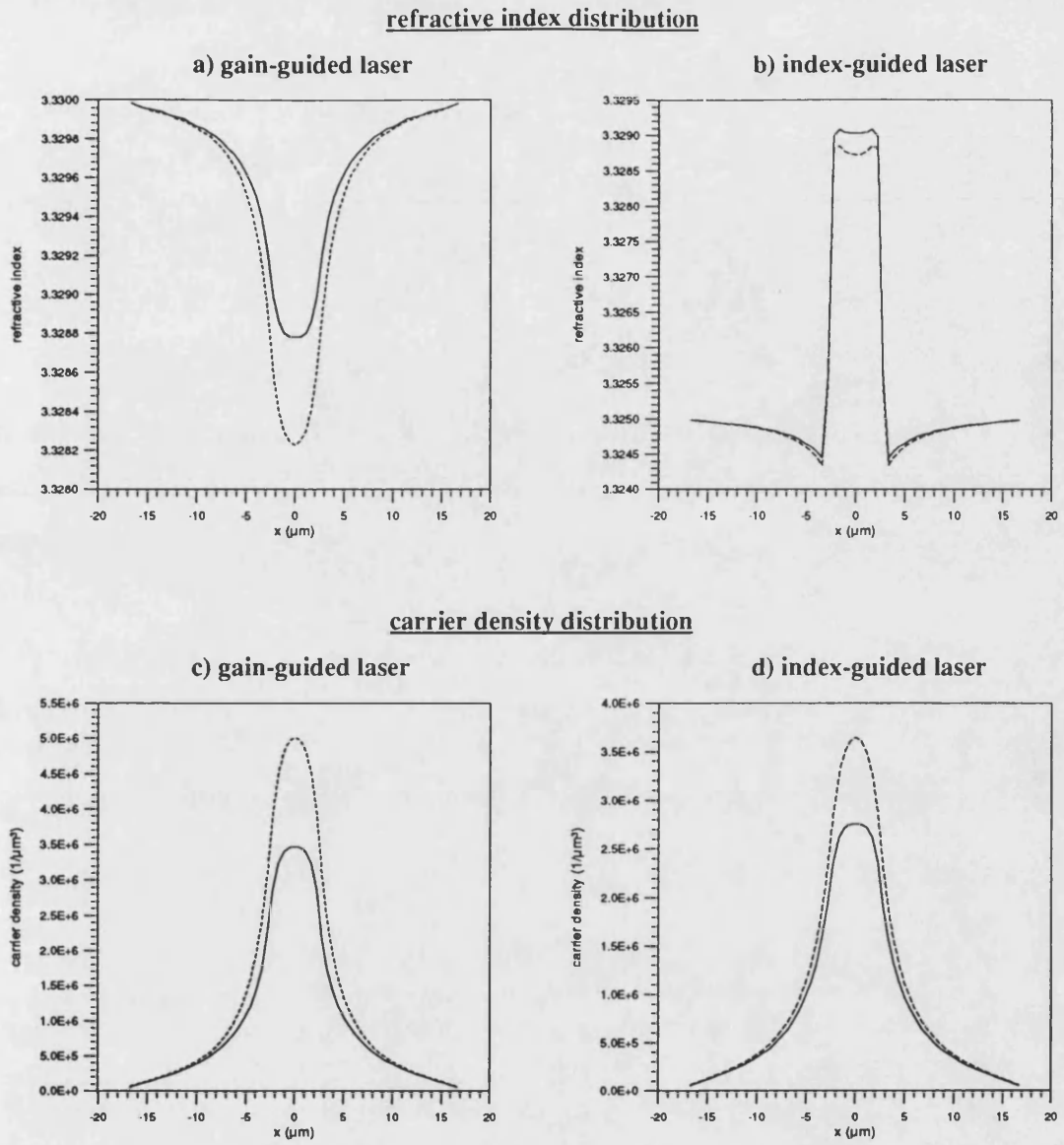


Fig. 7.5.4: Refractive index distribution and carrier density for gain-guided and index-guided laser: distribution before (broken line) and after (continuous line) self-consistency (J_o = 8.0 10^{-5} A/μm²).

5.1 Numerical aspects of the solution for gain-guided devices

The self-consistent solution for the gain-guided laser is found with the same iterative HGCM scheme used for the index-guided devices, as described in Section 2.5. However, a peculiar pattern is observed in the evolution of the near field with the iteration number in the modelling of gain-guided lasers, which is shown in Fig. 7.5.5.

Note in Fig. 7.5.5 that the field diffracts in the first iterations since there is no lateral confinement of the optical field ($\Delta n = 0$), and also because there is an anti-index-step produced by the injected carriers, equation (7.2.8). In the iteration process the field-carrier interaction takes place and self-consistently modifies the refractive index and the gain profile in the structure. Hence, when then the field-carrier interaction sets in and produces gain-guiding effects, the field 'recomposes' in the final mode-like shape, which is shown in Fig. 7.5.5 d) and also in Fig. 7.5.2.

However, it is noted that in the case of the gain-guided device the phase front at the facet, and consequently also the far field, cannot be calculated with the HGCM, perhaps because of numerical inaccuracies introduced by the process of 'recomposition' of the field, shown in Fig. 7.5.5.

Similar features are observed by varying the number of collocation points used in the HGCM; or by changing the initial field distribution in the iteration process, both a Gaussian and a supergaussian initial field distribution have been used obtaining the same self-consistent solution.

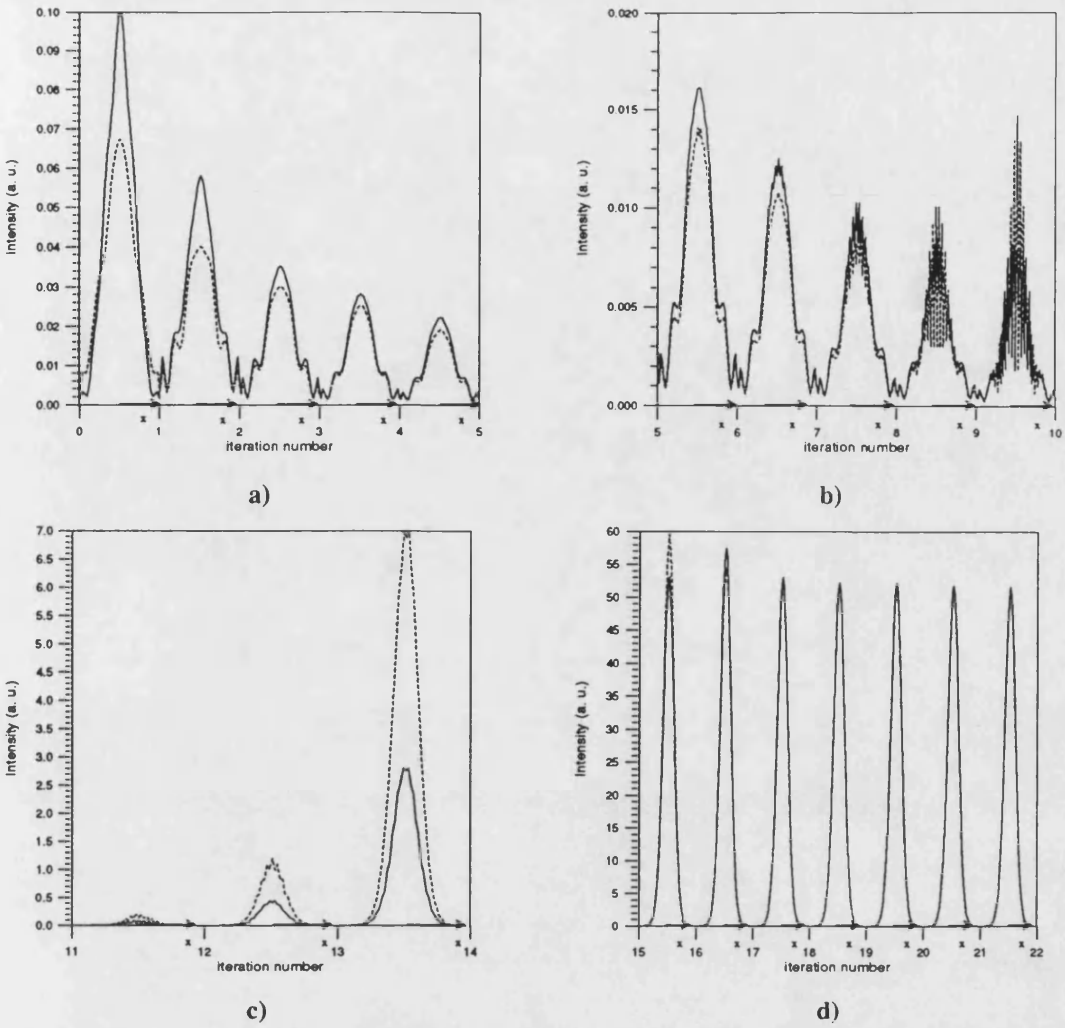


Fig. 7.5.5: Near field profile in the iterative process: broken line at z = 0; continuous line at z = L. Note that in a) and b) the scale is orders of magnitude smaller than that used in c) and d).

6. Modelling of active tapered devices

Two modelling techniques are next reviewed in the context of active devices, the Beam Propagation Method [Chapter 3, Section 2.2] and the mode matching method [Chapter 3, Section 2.1]. The purpose of the following discussion is to highlight the differences between the HGCM self-consistent scheme and the two methods reviewed in this section. To facilitate a direct comparison of the equations used in the two methods reviewed here and those presented in Section 2 of this Chapter, the same parameters will be used in the present section. Fig. 7.2.1 shows a schematic of a tapered device and defines the co-ordinate axes used in the following discussion.

6.1 Beam Propagation Method

The Beam Propagation Method (BPM) which is commonly used to solve for field propagation in (longitudinally non-uniform) structures, has been recently incorporated in self-consistent schemes for the modelling of tapered active devices, for example, [12], [23], [24]. In such computation schemes the optical field is propagated along the structure using the BPM algorithm [Chapter 3, Section 2.2], and the coupling between the optical field and the carrier density is ensured by varying the gain and the effective refractive index in the structure consistently with the optical field intensity.

In [23] the linearly tapered amplifier section of MOPA (Master Oscillator Power Amplifier) structures has been analysed [as in Section 3.1 of this Chapter]. The relevant parameters used in the BPM model presented in [23] are summarised below. The built-in and carrier and temperature dependent complex refractive index distribution in the tapered device is given by

$$n_{eff}(x,z) = n_{eff}^{(o)} + \frac{1}{2k_o}\left[\Delta n_R(x,z,N) + i\Gamma g(x,z) - i\alpha_{loss}\right] + \alpha_T T(x,z) \quad (7.6.1)$$

with $n_{eff}^{(o)}$ the unperturbed effective refractive index, Γ is the confinement factor along the vertical (y) axis, g(x,z) the local gain, $\Delta n_R(x,z,N)$ the change in refractive index due to the injected carriers [Section 2.2, this Chapter], α_{loss} the linear losses (including the scattering losses and the losses due to free carriers), α_T the temperature coefficient of effective index (in K^{-1}), and T(x, z) is the local temperature change. The saturated gain g(x, z) is as defined as in equation (7.2.11).

The model presented in [23] is particularly convenient to model gain-guided devices since the inclusion of a lateral index step to produce index-guiding effects generally introduces complications in the formalism.

In [24], [25] it is shown that when applying the self-consistent BPM model [23] to laser devices, chaotic behaviour may be observed. In [24], in fact, the same BPM model described in [23] forms the basis of a numerical Fox-Li technique in which the field is propagated between the laser facets (both with finite reflectivities) until it converges to a stationary field distribution. However, it is not clear if the instabilities discussed in [24], [25], reflect physical instabilities which have been observed experimentally in tapered devices because the BPM model in [24] uses steady-state equations. Depending on the driving current (which represents the control parameter) it is observed in [24] that the output near field and power do not always converge, but may show chaotic behaviour. It is also shown that the critical parameter that may lead to chaos in the modelling of flared lasers is the antiguiding factor, β_a in equation (7.2.10).

6.2 Mode matching method

Another commonly used method for computing field propagation is the mode matching method [Chapter 3, Section 2.1]. In the following discussion reference is made to [26] and [18] where the mode matching method has been applied to active devices with linear and nonlinear tapers. With the mode matching technique the tapered structure is approximated by a series of (longitudinally uniform) slab waveguide sections in each of which the local modes are determined. Considering the initial field to be the fundamental local mode at $z = 0$. The coupling between the initial field and the higher order and radiation modes supported by each uniform section is thus calculated matching the field components at the interface between adjacent sections. In the method presented in [26] self-consistency is ensured by coupling the diffusion equation to the equation that governs the amplification rate of the optical intensity along the tapered amplifier. The carrier diffusion equation used there is as in equation (7.2.14), while the field (intensity) equation is

$$\frac{dS}{dz} = \left(\frac{\Gamma g_u}{1 + \frac{S}{S_{sat}}} - \alpha_{loss} - \alpha_{tap} \right) S \qquad (7.6.2)$$

where $S(z) = \frac{1}{wd} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} S(x,y,z)dxdy$ is the optical field intensity averaged across the x-y plane, g_u is the unsaturated material gain, Γ the optical confinement factor in the x-y plane, S_{sat} the saturation intensity, α_{loss} the losses in the active and cladding layers, and $\alpha_{tap} = -w(z)\frac{d}{dz}\left(\frac{1}{w(z)}\right)$ where w(z) is the width of the local slab, the factor which accounts for the spreading of the optical intensity due to the tapered geometry. It is to be noted that in [26] the refractive index dependence on the carrier density is neglected since this effect is considered of secondary importance in index-guided devices (lateral refractive index step > 5 10^{-3}). This formalism has been used to analyse linearly [26] as well as exponentially tapered travelling-wave amplifiers [8].

Another example of self-consistent model based on the mode matching technique is described in [18]. As in [26] the tapered device is approximated by a series of slab waveguide sections. In [18], however, the optical field is expanded on the basis of the set of only the discrete (local) modes. Further, in [18] the carrier distribution is calculated by solving the carrier diffusion equation of the form (7.2.14) with the Jacobi-Tridiagonal method. Self-consistency is ensured by iteratively changing the gain and the refractive index distribution as the field propagates along the structure. Although similar to [26], the method described in [18] retains the refractive index dependence on carrier density, but only the propagation constants of the (local) modes are altered, leaving the modal field profiles unperturbed.

Both these models based on the mode matching technique, [26], [18] have been applied to index-guided structures for which the process of finding the local (bound) modes is relatively straightforward.

Summary

A comprehensive method based on the Hermite-Gauss expansion for achieving self-consistent models of semiconductor lasers and amplifiers has been presented in this chapter. The characteristic feature of this method is that it solves both the optical

field equations and the carrier diffusion equation with the same HGCM scheme. In comparison with other methods, the HGCM model offers the advantage of implicitly including the contribution of the radiation mode set. This is achieved by expanding the optical field in the structure on the basis of the complete set of Hermite-Gauss functions. The property of discreteness of the basis set simplifies the formalism of the present method since no integrals need to be included in the expansion, thereby reducing the complications deriving from the discretisation of the continuum spectrum (as with the mode matching method). By making use of the collocation method the HGCM is formulated in matrix form which, although convenient for the numerical solution of the problem, may in some cases also be a limiting factor for the accuracy achieved by the method.

In this chapter the overall iterative HGCM scheme has been tested with several examples. Gradual modifications on the self-consistent solution have been produced by individually varying the parameters used in the calculations. Some of the results obtained with the HGCM have been compared with those calculated with another method of solution based on the mode matching scheme (using only the discrete set of modes). The comparisons are satisfactory, although further investigations are needed to fully assess the capabilities of the HGCM. However, the variety of devices analysed in this chapter proves that the HGCM scheme is convenient and readily applicable for the analysis of a category of longitudinally non-uniform active optical devices. The HGCM has been applied also for analysing gain-guided narrow stripe lasers. It is argued that being a total field analysis, the HGCM can be applied even in the extreme case of a purely gain-guided device where the local mode expansion scheme is not as convenient because of the difficulties arising from the consideration of the radiation mode set.

Appendix 7.1

Material parameters of typical semiconductor devices

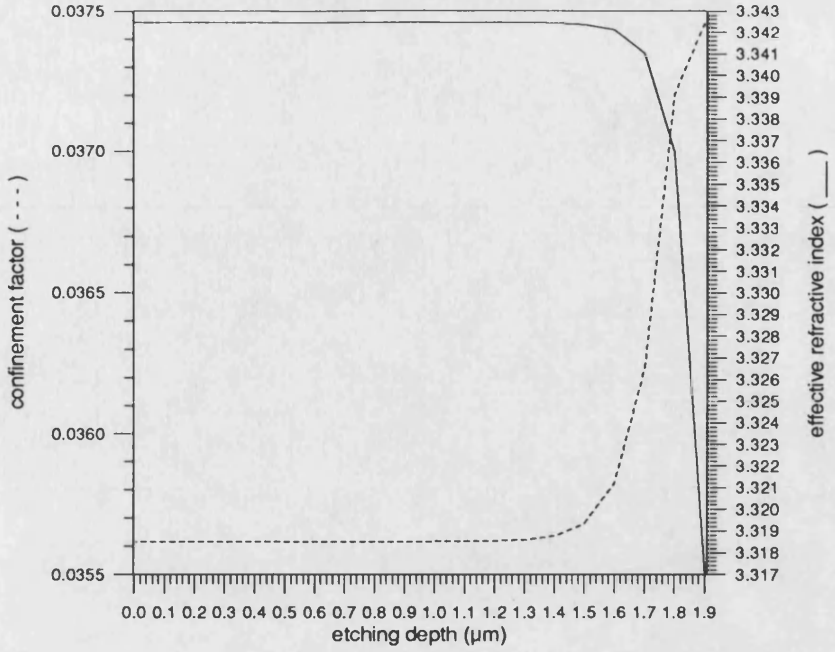
The multilayer structure of the material from which typical semiconductor devices are made is described in Tab. A7.1-1 for Quantum Well material and Tab. A7.1-2 for bulk material, [27]. Most of the devices that have been measured are characterised by a rib to obtain (weak) guiding also along the transverse (x) axis. The rib is produced by reaction ion beam etching, and hence it is useful to be able to predict the effect of the etching depth on the (calculated) effective refractive index in the structure, Fig. A7.1-1 a) and b).

Tab. A7.1-1: Layer structure of a typical Quantum Well material (QT829A)

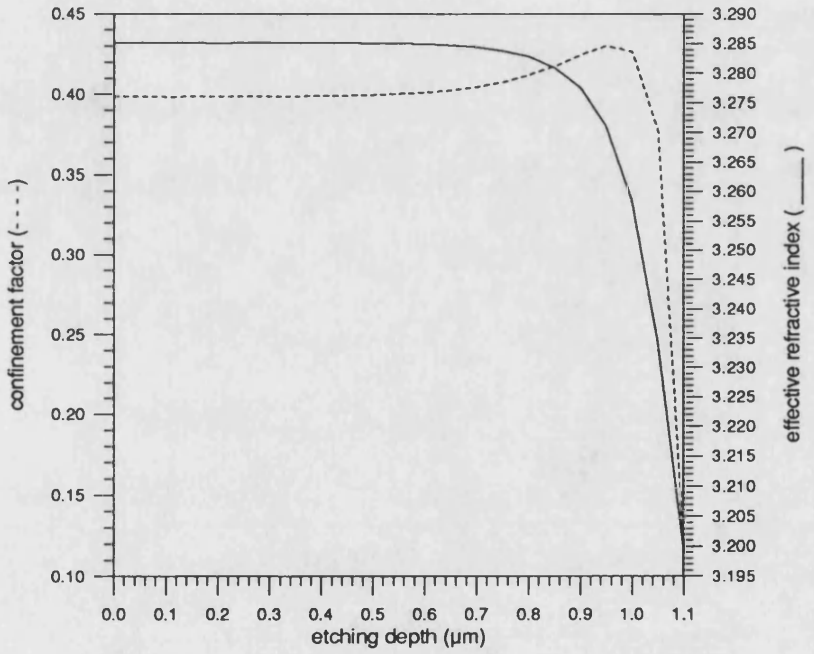
material	thickness	doping
GaAs	0.3µm	$9.0\ 10^{18}$ Zn
$Al_{0.60}$	1.4µm	$2.2\ 10^{18}$ C
$Al_{0.60}$	0.2µm	$\sim 5.0\ 10^{17}$ C
$Al_{0.26}$	0.15µm	undoped
GaAs	103Å	undoped
$Al_{0.26}$	100Å	undoped
GaAs	103Å	undoped
$Al_{0.26}$	0.15µm	undoped
$Al_{0.60}$	1.6µm	$1.3\ 10^{18}$ Si
GaAs	0.1µm	$1.0\ 10^{18}$ Si

Tab. A7.1-2: Layer structure for a bulk material (CB539)

layer	thickness (µm)	composition	doping type
contact	0.2	GaAs	p+
cladding	0.1	$Al_{0.60}Ga_{0.40}As$	p–
active	0.1	$Al_{0.04}Ga_{0.96}As$	u
cladding	2.0	$Al_{0.60}Ga_{0.40}As$	n–
substrate	- - -	GaAs (3° off axis)	n+



a)



b)

Fig. A7.1-1: Variation of the (calculated) effective refractive index and of the confinement factor along the vertical (y) direction, with respect to the etching depth; a) Quantum Well material (QT829A), b) bulk material (CB539).

Appendix 7.2

Average photon density

The objective in this Appendix is to provide the link between the electromagnetic fields in the structure and the photon density distribution that is used in the diffusion equation (7.2.14).

To be specific, let the typical device structure be that of a multilayer ridge waveguide, of width $2a_o$, Fig. A7.1-1, and it is assumed that the dominant guiding is along y.

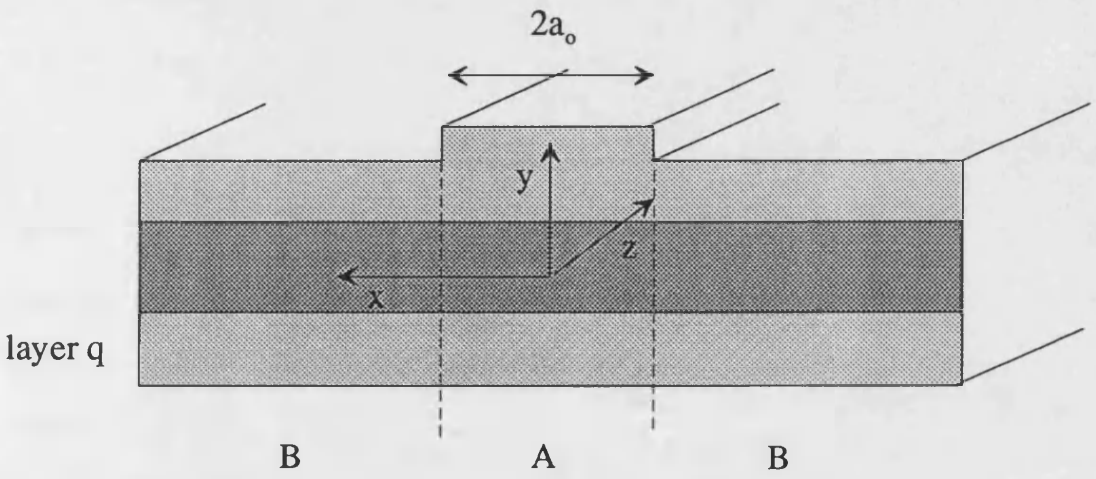


Fig. A7.2-1: Schematic of a typical ridge waveguide semiconductor device; $2a_o$ is the ribwidth.

The waveguide is defined by the complex refractive index distribution n(x,y,z), but for weakly non-uniform structures a scalar field analysis is acceptable. Because of the assumption of large aspect ratio the structure may be taken to support $(TE)_y$ and $(TM)_y$ modes. Consider, here, a $(TE)_y$-type of electromagnetic field characterised by the field components E_x, H_y, H_z. [Significant deviations from the large aspect ratio/Effective Dielectric Constant approximations are discussed, for example in [28], but will not be considered in the present formulation.]

With this approximation let $E_x(x,y,z)$ be the generating field component and use

$$F(x,y,z) = E_x(x,y,z) \qquad \text{(A7.2 - 1)}$$

F(x,y,z) satisfies the wave equation (1.4.3)

$$\left[\partial_x^2 + \partial_y^2 + \partial_z^2 + k_o^2 n^2(x,y,z)\right]F(x,y,z) = 0 \qquad \text{(A7.2 - 2)}$$

where n(x,y,z) is the complex dielectric distribution which accounts for gain/loss in the layers.

It is assumed that the Effective Dielectric Constant (E. D. C.) analysis is applicable such that $\left|\frac{\partial E_x}{\partial y}\right| >> \left|\frac{\partial E_x}{\partial x}\right|$, [Appendix 1.1]; hence, consider, for example, only the lateral (x) region A defined as $-a_o \le x \le a_o$, [for tapered devices, $a_o = a_o(z)$]. In region A the variable x is considered as a parameter, hence $F(x,y,z) = F_A(y,z)$ and satisfies

$$\left[\partial_y^2 + \partial_z^2 + k_o^2 n_A^2(y)\right] F_A(y,z) = 0 \qquad \text{(A7.2 - 3)}$$

with $n_A = n_{A,R} + i n_{A,I}$. The approximate solution of equation (A7.2 - 3) produces the effective refractive index

$$n_{eff,A} = \sum_q \Gamma_{A,q}\left(n_{A,q} + i\frac{g_{A,q}}{2k_o}\right) \qquad \text{(A7.2 - 4)}$$

where $\Gamma_{A,q}$ and $g_{A,q}$ are the confinement factor and material gain, respectively, in each layer (q) in region A for the vertical mode, $F_A(x,z) \approx f_A(y)e^{-ik_o n_{eff,A} z}$, with $f_A(y)$ calculated with the approximation that the layers have purely real refractive indices. Similarly for region B. [The effective refractive indices $n_{eff,A}$ and $n_{eff,B}$ are used in the two-dimensional HGCM analysis, [Chapter 5, Section 2.5], as n_1 and n_2, respectively.]

Then, on the basis of the E. D. C. analysis

$$F(x,y,z) = F_h(x,z)f_v(y) \approx F_h(x,z)f_A(y) \approx F_h(x,z)f_B(y) \qquad \text{(A7.2 - 5)}$$

with $f_A(y)$ and $f_B(y)$the modal profiles of the vertical fields, as defined in (A7.2 - 3).

The field $F_h(x,z)$ thus satisfies the wave equation (1.4.4)

$$\left[\partial_x^2 + \partial_z^2 + k_o^2 n_{eff}^2(x,z)\right] F_h(x,z) = 0 \qquad \text{(A7.2 - 6)}$$

where $n_{eff} = \begin{cases} n_{eff,A} & |x| \le a_o \\ n_{eff,B} & |x| > a_o \end{cases}$.

Also assume

$$F_h(x,z) \approx \tilde{f}_h(x,z)e^{-ipz} \qquad \text{(A7.2 - 7)}$$

where $\tilde{f}_h(x,z)$ is the slowly varying field amplitude, and p a positive, real constant; typically p is chosen to be in the range $k_o n_{eff,B} \le p \le k_o n_{eff,A}$.

From equation (A7.2 - 1),

$$E_x(x,y,z) \approx f_v(y)\tilde{f}_h(x,z)e^{-ipz} \qquad \text{(A7.2 - 8)}$$

The magnetic field components derive from Maxwell' s equations [Chapter 1, Section 3.1 and Appendix 1.4], in particular the transverse component of the magnetic field is

$$H_y = \frac{i}{\omega\mu_o}\partial_z E_x \qquad \text{(A7.2 - 9)}$$

Assuming that as a function of z, $\tilde{f}_h(x,z)$ varies slowly compared to e^{-ipz}, then equation (A7.2 - 9) may be reduced to

$$H_y \approx \frac{p}{\omega\mu_o}E_x \qquad \text{(A7.2 - 10)}$$

It is convenient to write the electric field (A7.2 - 8) as

$$F(x,y,z) = \frac{A}{c_h c_v}f_h(x,z)f_v(y) \qquad \text{(A7.2 - 11)}$$

where $f_h(x,z) = \tilde{f}_h(x,z)e^{-ipz}$ and $f_v(y)$ are non-normalised functions, $\frac{A}{c_h c_v}$ a normalising constant and $c_h^2 = \int_{-\infty}^{+\infty}|f_h(x,z=0)|^2 dx$ and $c_v^2 = \int_{-\infty}^{+\infty}|f_v(y)|^2 dy$ [refer to Appendix 1.4]. For simplicity, choose $F_h(x,z) = \frac{f_h(x,z)}{c_h}$ and $F_v(y) = \frac{f_v(y)}{c_v}$ (both in $\sqrt{\frac{V}{m}}$, while A is dimensionless), so that, at any z,

$$F(x,y,z) = AF_h(x,z)F_v(y) \qquad \text{(A7.2 - 12)}$$

Thus, at any z, the power density associated with F(x, y, z) is

$$S_z(x,y,z) = A^2\frac{p}{2\omega\mu_o}|F_h(x,z)|^2|F_v(y)|^2 \qquad \text{(A7.2 - 13)}$$

and in this case the total power is

$$S_z = A^2\frac{p}{2\omega\mu_o}\left(\int_{-\infty}^{+\infty}|F_h(x,z)|^2 dx\right)\left(\int_{-\infty}^{+\infty}|F_v(y)|^2 dy\right) \qquad \text{(A7.2 - 14)}$$

Note that $\int_{-\infty}^{+\infty} |F_h(x,z)|^2 dx$ is > (<) 1 in active gain (loss) devices. Hence, the total power associated with F(x, y, z) becomes

$$S_z = A^2 \frac{p}{2\omega\mu_o} \int_{-\infty}^{+\infty} |F_h(x,z)|^2 dx \qquad \text{(A7.2 - 15)}$$

At z = 0 the power carried by the initial field F(x, y, z = 0) (which can be interpreted as the initial guess of the iterative process in the laser model) is known, P_o, and it is

$$P_o = A^2 \frac{p}{2\omega\mu_o} \qquad \text{(A7.2 - 16)}$$

Thus, the number of photons crossing the unit area in the unit time interval is

$$P(x,y,z) = \frac{S(x,y,z)}{h\nu} = \frac{P_o}{h\nu} |F_h(x,z)|^2 |F_v(y)|^2 \qquad \text{(A7.2 - 17)}$$

and the number of photons crossing the unit area in the unit time interval, averaged over the thickness (d) of the active layer, is

$$P_{av}(x,z) = \frac{P_o}{h\nu} |F_h(x,z)|^2 \frac{1}{d} \int_{-\frac{d}{2}}^{\frac{d}{2}} |F_v(y)|^2 dy = \frac{\Gamma}{d} \frac{P_o}{h\nu} |F_h(x,z)|^2 \qquad \text{(A7.2 - 18)}$$

where Γ is the vertical confinement factor

The stimulated emission term in the diffusion equation (7.2.14) thus takes the form

$$g(x,z) P_{av}(x,z) = \frac{\Gamma P_o}{dh\nu} g(x,z) |F_h(x,z)|^2 \qquad \text{(A7.2 - 20)}$$

with g(x,z) the local gain.

References

[1] A. V. Chelnokov, J. M. Louritioz, P. Gavrilovic, *Numerical modeling of the spatial and spectro-temporal behavior of wide-aperture unstable resonator semiconductor lasers*, IEEE Photonics Technology Letters, 7 (1995), pp. 863-865

[2] J. V. Moloney, R. A. Indik, C. Z. Ning, *Full space-time simulation for high-brightness semiconductor lasers*, IEEE Photonics Technology Letters, 9 (1997), pp. 731-733

[3] Y.Champagne, S. Mailhot, N. McCarthy, *Numerical procedure for the lateral-mode analysis of broad area semiconductor lasers with an external cavity*, IEEE Journal of Quantum Electronics, 31 (1995), pp. 795-810

[4] H. Adachihara, O. Hess, E. Abraham, P. Ru, J. V. Moloney, *Spatiotemporal chaos in broad-area semiconductor lasers*, Journal of the Optical Society of America B, 10 (1993), pp. 658-665

[5] R. J. Lang, A. G. Larsson, J. G. Cody, *Lateral modes of broad area semiconductor lasers: theory and experiment*, IEEE Journal of Quantum Electronics, 27 (1991), pp. 312-320

[6] D. Mahuys, R. J. Lang, M. Mittlestein, J. Salzman, A. Yariv, *Self-stabilised nonlinear lateral modes of broad area lasers*, IEEE Journal of Quantum Electronics, QE-23 (1987), pp. 1909-1920

[7] G. B. H. Thompson, *A theory for filamentation in semiconductor lasers including the dependence of dielectric constant on injected carrier density*, Optoelectronics, 4 (1972), pp. 257-310

[8] G. Bendelli, K. Komori, S. Arai, and Y. Suematsu, *A new structure for high-power TW-SLA*, IEEE Photonics Technology Letters, 3 (1991), pp. 42-45

[9] D. F. Welch, R. Parke, D. Mehuys, A. Hardy, R. Lang and D. Scifres, *1.1 W CW diffraction-limited operation of monolitically integrated flared-amplifier master oscillator power amplifier*, Electronics Letters, 28 (1992), pp. 2011-2013

[10] K. A. Williams, J. Sarma, I. H. White, R. V. Penty, I. Middlemast, T. Ryan, F. R. Laughton and J. S. Roberts, *Q-switched bow-tie lasers for high-energy picosecond pulse generation*, Electronics Letters, 30 (1994), pp. 320-321

[11] C. Vassallo, *Analysis of tapered mode transformersfor semiconductor optical amplifiers*, Optical and Quantum Electronics, 26 (1994), pp. S235-S248

[12] P. Chalzan, J. D. Ralston, *Beam propagation model of tapered amplifiers including non-linear gain and carrier diffusion*, Digest of the conference on Semiconductor Lasers, Advanced Devices and Applications, vol. 20, Keystone, Co (U.S.A.), August 1995, pp. 92-94

[13] J. N. Walpole, *Semiconductor amplifiers and lasers with tapered gain regions*, Optical and Quantum Electronics, 28 (1996), pp. 623-645

[14] G. P. Agrawal and N. K. Dutta, *Long-Wavelength Semiconductor Lasers*, Van Nostrand Reinhold (1986)

[15] J. Huang, L. W. Casperson, *Gain and saturation in semiconductor lasers*, Optical and Quantum Electronics, 25 (1993), pp. 369-390

[16] G. Fox, T. Li, *Resonant Modes in a Maser Interferometer*, Bell System Techonoly Journal, 40, (1961), pp. 453-458

[17] I. Middlemast, J. Sarma, P. S. Spencer, *Characteristics of tapered rib-waveguides for high power semiconductor optical sources*, IEE Proceedings Optoelectronics, 144 (1997), n.1, pp. 8-13

[18] N. Brooks, J. Sarma, I. Middlemast, *A compact model of tapered geometry semiconductor optical devices*, Semiconductor and Integrated Optoelectronic Conference (SIOE '96), paper 45, abstract, Technical Digest, Cardiff, April 1996

[19] N. S. Brooks, J. Sarma, I. Middlemast, *A new design for tapered-geometry high-power semiconductor optical sources*, in proceedings LEOS '96, paper n. WZ5.

[20] A. F. Milton, W. K. Burns, *Mode coupling in optical waveguide horns*, IEEE Journal of Quantum Electronics, QE-13 (1977), pp. 828-835

[21] G. P. Agrawal, *Lateral analysis of quasi-index-guided injection lasers: transition from gain to index guiding*, Journal of Lightwave Technology, LT-2 (1984), pp. 537-543

[22] C. L. Reynolds, W. R. Holbrook, J. A. Shimer, S. M. Tharaldsen, G. P. Agrawal, H. Temkin, *Experimental verification of transition from gain- to index-guiding in a rib-waveguide AlGaAs laser*, Electronics Letters, 22 (1986), pp. 1290-1291

[23] R. J. Lang, A. Hardy, R. Parke, D. Mehuys, S. O'Brien, J. Major and D. Welch, *Numerical analysis of flared semiconductor laser amplifiers*, IEEE Journal of Quantum Electronics, 29 (1993), pp. 2044-2051

[24] G. Levy, A. A. Hardy, *Control and suppression of chaos in flared laser systems: a numerical analysis*, IEEE Journal of Quantum Electronics, 34 (1998), pp. 1-6

[25] G. Levy, A. A. Hardy, *Chaotic effects in flared lasers: a numerical analysis*, IEEE Journal of Quantum Electronics, 33 (1997), pp. 26-32

[26] G. Bendelli, K. Komori, S. Arai, *Gain saturation and propagation characteristics of index-guided tapered-waveguide travelling-wave semiconductor laser amplifiers (TTW-SLAs)*, IEEE Journal of Quantum Electronics, 28 (1992), pp. 447-457

[27] T. Ryan, internal communication

[28] L. G. Ferreira, M. A. A. Pudensi, *Waveguiding in a dielectric medium varying slowly in one transverse direction*, Journal of the Optical Society of Amerca, 71 (1981), pp. 1377-1380

Future work

Reflected field: The analysis of longitudinally non-uniform devices is based on the paraxial approximation which is quite valid for weakly non-uniform structures, such as adiabatic tapers. For continuously non-uniform structures the 'forward' and 'reverse' travelling components do not strictly decouple even in the paraxial approximation; for the present analysis, however, a decoupled formalism is used. Note though that although such an approximation may be satisfactory with weakly non-uniform, passive devices, in active devices the small, but continuous, reflection could, in some cases, considerably affect the device characteristics.

Complex w_o: The analysis of structures defined by complex refractive index distributions has also been analysed with the HGCM and satisfactorily accurate solutions have been mostly obtained. Nevertheless, the solution of complex eigenvalue problems should be properly solved using a complex width parameter w_o for the HG expansion functions. However, the complications introduced by such a modification have been considered to be substantial, and hence they have been deferred to future developments.

Gain-guided devices: Further investigations are also needed in the modelling of gain-guided devices. In this respect it has been possible to obtain results with the HGCM for gain-guided stripe lasers that are in good qualitative agreement with those found in the literature. However, it is felt that some numerical aspects, e.g., pertaining to the complex mode functions, need to be better understood to establish the potential of the HGCM self-consistent scheme in this area.

Vectorial analysis: The present HGCM formulation relies on the scalar analysis of the field since the structures of interest are weakly guiding. But it is felt that the analysis should be extended to include vectorial aspects such as the discrimination between TE and TM mode solutions, and, for example, in longitudinally non-uniform devices, the full vectorial characterisation of the Poynting vector.

Time-dependent effects: In this thesis the HGCM self-consistent model has been presented in the steady-state formulation. However, the possibility of including time-dependent effects in the form of a slowly varying time-dependence must be considered. The time dependence could be included in different stages, the first of which would be to simply introduce slow time-dependence in the diffusion equation using a finite difference approximation. At a later stage the solution of the field equation could be extended to include faster time-dependent effects.

Laguerre-Gauss basis functions: The solution scheme presented in this thesis may be effectively extended to the analysis of cylindrical structures by replacing the Hermite-Gauss with the Laguerre-Gauss function set. The possibility of also solving the carrier diffusion equation in cylindrical co-ordinates should be addressed. This would permit a self-consistent model applicable, e.g., to Vertical Cavity Surface Emitting Lasers.

As is always the case, the detailed study of a new topic 'answers' many 'questions', but raises many others. It is in that sense that the points outlined under Future Work should be read. However, that should not distract from the fact that the HGCM developed in the course of this research and presented in this thesis has been shown to be a new and very effective method for modelling a wide range of optical devices.

Conclusions

The objective of the research presented in this thesis has been to develop a new procedure for analysing active, semiconductor optical devices, particularly devices which are longitudinally non-uniform. The development of the proposed self-consistent scheme has progressed in three different stages. The first has been to formulate and test the method for the electromagnetic field analysis in such devices. In the second stage it has been shown that with appropriate modifications, essentially the same procedure that has been developed for the optical field analysis may also be applied to very effectively solve the carrier diffusion equation typically associated to active semiconductor devices. The final stage of the work has been to combine the schemes developed in the first two stages in an appropriately iterative procedure to achieve the desired model which self-consistently solves the carrier density and the photon field equation in active semiconductor optical devices such as lasers and optical amplifiers.

Hermite-Gauss functional form: The first approach to the analysis of electromagnetic field problems has been to study the properties pertinent to two important sets of functions: the Hermite-Gauss Beams (HGBs) and the Hermite-Gauss Eigenfunctions (HGEs). Although these two sets have been used extensively in the literature, the review discussed in this thesis has been useful not only in collecting together various concepts and features pertaining to these functions, but, importantly, to also clearly identify the differences between the two sets of functions. In the present context it is observed that i) both the HGBs and the HGEs provide a complete set of orthogonal functions in the same functional space of square integrable functions; and ii) the HGBs are most often used in the analysis of diffracting field, while the HGEs are far more convenient for the analysis of field propagation in inhomogeneous media.

HGEs as basis functions: The basic point to note is that the optical devices to be analysed are essentially dielectric structures with deviations from typical waveguides due to longitudinal non-uniformities. Therefore the analysis proposed

here is to express the (total) field in these structures as an expansion in terms of a suitable, complete set of orthogonal functions. Since the media are in general inhomogeneous, the arguments from the immediately previous paragraph strongly suggest that the HGEs functions should be used as the complete set of basis functions. Other features that add to the justification for this choice are:

i) individually the HGEs closely resemble corresponding bound modes of symmetric slab dielectric waveguides, hence in many cases the total field in the device can be expected to be very well represented by only a few HGEs;

ii) because the HGEs form a complete set for the pertinent functional space of interest the contribution of the continuum (radiation) modes is taken into account by the HG field expansion;

iii) although an infinite number of HGEs is needed for the correct expansion, the set is discrete so that no integral representation is needed (in contrast with the expansion using dielectric slab modes);

iv) the HGEs individually satisfy the radiation condition at infinity, and hence that condition is built into the field expansion.

HGCM - field analysis: Having chosen the HGEs as the basis functions, the analytic expansion procedure has been developed to obtain a set of coupled integro-differential equations for evaluating the expansion coefficients. From a computational viewpoint integrations are very laborious, and hence the more efficient numerical collocation method is adopted to overcome such difficulties. Thus, the Hermite-Gauss Collocation Method (HGCM) forms the chosen scheme in this thesis for analysing passive and active optical devices, both longitudinally uniform and non-uniform.

The HGCM has been extensively tested for a variety of field problems, such as i) longitudinally uniform waveguides, ii) diffraction in half-space, iii) abrupt discontinuities in dielectric waveguides, iv) coupled waveguides, v) tapered devices.

Whenever possible these solutions have been compared with corresponding analytic results, or with results from alternative numerical methods. In all cases the HGCM results closely match with the others.

A point to note is that in the present formalism the 'forward' and 'reverse' travelling wave components are assumed to be decoupled even though, strictly, such decoupling is not possible for continuously non-uniform structures.

The computation time is one measure of efficiency of any model. The HGCM is found to use more computation time than the Discrete Mode Matching (DMM) method - this being the only readily available numerical method against which comparison could be made. But the comparison is not on the same grounds since the DMM uses only the discrete modes of the local waveguide as the expansion functions - the continuum modes are ignored but their absence poses serious limitations to the DMM in many cases while including them would vastly increase computation time. The HGCM, on the other hand, implicitly includes, in effect, the continuum spectrum.

The w_o parameter: In the expansion of an arbitrary function the infinite set of HGEs has to be used, in which case any value of w_o will, in principle, be correct. Practical computations, however, use only a finite number of terms in the series expansion, which poses the problem of choosing an appropriate w_o for a sufficiently accurate representation. This interesting issue has been discussed in this thesis. It is observed that the number of expansion terms that are typically used in the calculations is very large (in order to obtain a detailed description of the field profile) and hence the HGCM results are not significantly affected by the choice of w_o.

Nevertheless, in the early stages of the work it was found that the solutions computed with the HGCM did quite strongly depend on w_o (even with a large number of expansion terms). This 'problem' was identified as due not to the truncated expansion, but to the fact that the computation of step index profile waveguides with the HGCM led to sampling errors. It was found that only a few particular values of w_o for that case gave accurate results, although experience and educated guesses helped to obtain the appropriate w_o quite rapidly.

In a novel development it has been shown that by replacing the step index profile with a sharp but analytic profile, defined by a supergaussian function, completely eliminates the 'problem'. Thus it has been found that, in this case, even using a relatively small number of expansion terms yields very accurate results. This has proved to be a very important innovation.

HGCM - carrier diffusion equation: A particularly useful feature of this research has been the development of a new numerical procedure for the solution of the carrier diffusion equation that is typical of semiconductor optical devices. It was proposed and demonstrated within this research programme that the HGCM (suitably modified) provides a very effective means for solving the (linear and nonlinear) carrier diffusion equation. Results calculated with the HGCM have been compared with analytic solutions, or with other numerical solutions, obtaining most favourable comparisons. Particularly satisfying is the fact that the HGCM works equally effectively with relatively narrow and very broad current injection regions which makes it very amenable for use in the modelling of taper geometry devices.

HGCM self-consistent scheme: The two HGCM schemes developed in the first part of the research work, the first to solve the (electromagnetic field) scalar wave equation, the second for the nonlinear (carrier) diffusion equation, have been merged in a self-consistent iterative scheme for the (steady state) modelling of active semiconductor optical devices.

A set of results obtained with the HGCM for linearly tapered active devices has also been used as a form of test for the HGCM by putting to zero the influence of some of the parameters at a time in attempts to understand computed results with expectations from experience. Further results for flared amplifiers and for parabolic taper lasers have been compared with those obtained from the approximate Discrete Mode Matching (DMM) method. The two methods agree satisfactorily well where the continuum (radiation) modes are not significant, but otherwise the two sets of results agree less precisely. This lack of agreement is because the DMM used here ignores radiation modes in the formulation.

Note, however, that it has not been possible to compare the test results obtained with the HGCM for linearly tapered lasers with those form the DMM since the latter becomes quite unacceptable as a model (without the inclusion of the continuum modes) for this device geometry. This shortcoming of the DMM is far less detrimental in the modelling of parabolic structures because in these devices the conversion to higher order (local) modes is relatively small.

List of Tables

5.1.1 Effective refractive index for waveguide bound modes calculated with HGCM, TMM and analytic solution (A) 105

5.1.2 Effective refractive index of the (bound) modes of three-layer slab waveguides with complex dielectric distribution in the core layer 107

5.2.1 Comparison of the computing time needed for HGCM and DMM to solve initial value problems. In particular the CPU time (minutes: seconds) is given as a function of M for the HGCM 127

6.3.1 Summary of the CPU time (in s) needed by the HGCM and other methods of solution - the JTM, the perturbation method (PM), and the transfer matrix method (TMM). 142

7.3.1 Parameters defining the tapered geometry structure of Fig. 7.2.1 c) . 165

7.3.2 Parameters used for modelling the tapered geometry amplifier . 167

7.5.1 Parameters used for modelling stripe lasers 189

A7.1-1 Layer structure of a typical Quantum Well material (QT829A) . 199

A7.1-2 Layer structure for a bulk material (CB539) 199

List of Figures

1.2.1 Interface between two homogeneous media defined by ε_1 and ε_2; $\underline{n}$ and $\underline{t}$ are the unity vectors perpendicular and parallel to the interface, respectively. . 5

1.7.1 Connections between the solutions of the paraxial wave equations (1.7.9) and (1.7.10). 14

1.9.1 Parallel plate metal closed waveguide: perfectly conducting boundaries at $x = \pm a$ 15

1.10.1 Schematic of a longitudinally uniform three-layer symmetric slab waveguide, the core layer (shaded region) has a larger dielectric constant with respect to the cladding layers. The thickness of the waveguide core layer is 2a . 18

1.10.2 Spectrum of the eigenvalues of an open waveguide (β_r and β_i refer to real and imaginary part of β respectively, n_1 is the highest value of the refractive index distribution, n_2 the lowest). 19

A1.1-1 Typical semiconductor device etched from multilayer material. In this specific example the variation of the thickness of the layers produces the refractive index step. 24

A1.1-2 Two-dimensional equivalent structure of the one shown in Fig. A1.1-1 obtained using the E. D. C. method. 25

A1.2-1 Decomposition on the wave vector $\underline{p}$ ($p = k_o\varepsilon_h$) along the transverse (x) and longitudinal (z) axes. 26

2.1.1 Diffraction of the fundamental Gaussian Beam (propagation is assumed to be only in the +z-direction). The broken curves show the phase front which is flat

at the beam waist ($w = w_o$) and curved at $z > 0$, and also the beam waist increasing with z. 40

2.2.1 The first three Hermite-Gauss Eigenfunctions. 45

3.2.1 Tapered structure approximated by a series of waveguide slabs (e.g., sections p and q). Input and output widths: w_o and w_f, respectively, length of the device: L. 60

4.5.1 Top view of a taper structure device: $2a_o$ and $2a_{fin}$ are the narrower and wider width, respectively, L is the length of the device. . . . 83

4.5.2 Longitudinal sections, characterised by different values for the width parameter, e.g., w_1 and w_2, for the analysis of tapers which have significantly varying widths. 83

5.1.1 Effective refractive index as a function of the width parameter, w_o, used in the HGCM calculations. The multimoded waveguide is characterised by the following parameters: $2a_o = 3\mu m$, $\lambda_o = 0.86\mu m$, $n_1 = 3.33$, $n_2 = 3.32$, $n_{eff(o)} = 3.32834$, $n_{eff(1)} = 3.32376$ 99

5.1.2 Refractive index distribution ($n_1 = 3.33$, $n_2 = 3.32$, $2a_o = 3\mu m$): solid line: supergaussian ($q = 16$); broken line: piecewise constant . . 100

5.1.3 Effective refractive index as a function of the w_o parameter used in the HGCM calculations using a continuous SG refractive index distribution ($q = 16$). (Parameters in Fig. 5.1.1) 101

5.1.4 Effective refractive index as a function of the w_o parameter and of the number of expansion term used in the HGCM calculations: a) $M = 11$, b) $M = 41$. The curves refer to broken line ST refractive index distribution, solid line SG refractive index distribution ($q = 16$). (Parameters in Fig. 5.1.1) . 103

5.2.1 Diffraction in half-space: the initial Gaussian field is specified at z = 0, z_{fin} is the propagation length, w(z = 0) = initial beam width, w(z) = beam width at the longitudinal position z. 108

5.2.2 Diffraction in half-space (λ_o = 1.55μm, the beam waist of the initial field distribution is w(z = 0) = 2.5μm, the refractive index of the homogeneous medium is n_h = 3.42, z_{fin} = 500μm): dotted line: Gaussian initial field distribution; solid line: HGCM (M = 99, w_o = 2.8μm); x: fundamental Gaussian Beam at z = z_{fin}; broken line: Plane Wave Decomposition method 109

5.2.3 Propagation in a multimoded waveguide (λ_o = 0.86 μm, n_1 = 3.33, n_2 = 3.32, $2a_o$ = 4μm, $n_{eff(o)}$ = 3.3289, $n_{eff(1)}$ = 3.3258, $n_{eff(2)}$ = 3.3213, respectively for each bound mode). The results are labelled as follows: dotted line: input field, solid line: HGCM (M = 99, w_o = 0.776μm, z_{fin} = 100μm), (x): analytic solution. The rib is shown as a rectangle 110

5.2.4 Propagation in a multimoded waveguide: contribution of the HG expansion functions to the field calculated in Fig. 5.2.3. (o): coefficients of the HGCM solution, (+): coefficients of the analytic solution. Mote that although M = 99 only the coefficients of the HG functions of order < 20 are shown here since the contributions of the HG functions of order > 20 are negligible . 111

5.2.5 Coupled waveguides (λ_o = 1.55 μm, n_1 = 3.28448, n_2 = 3.28241, $2a_1 = 2a_2$ = 3μm, d = 3μm). The input field is launched in one of the waveguides. The coupling length is $L = 2\pi/\Delta\beta \approx 2240\mu m$ since the propagation constants of the fundamental and first order modes of the structure are $\beta = k_o n_{eff}$, with $n_{eff(o)}$ = 3.28334 and $n_{eff(1)}$ = 3.28274. 112

5.2.6 Field propagation in coupled waveguides (z_{fin} = 1120μm, other parameters as shown in Fig. 5.2.5): solid line: HGCM (M = 99, w_o = 2.0μm), x: analytic solution, broken line: initial field distribution . . . 113

5.2.7 Junction between two waveguides, WG1 and WG2, of width $2a_1$ and $2a_2$, respectively. 114

5.2.8 Field propagation across a step discontinuity calculated at distance of 50μm from the junction: dotted line: input field, solid line: HGCM (M = 99, w_o = 1.2 μm), broken line: PWAS method. Parameters: $2a_1$ = 6μm, $2a_2$ = 2μm, n_1 = 3.32, n_2 = 3.3, λ_o=1.55μm. The mode effective refractive indices are, in WG1: $n_{eff(o)}$ = 3.31842, $n_{eff(1)}$=3.31378, $n_{eff(2)}$ = 3.30652, in WG2: $n_{eff(o)}$ = 3.31194 116-117

5.2.9 Top view of a linearly tapered structure. 118

5.2.10 Optical field propagating in a tapered structure, the inuput field is the (local) fundamental mode at z =0 (n_1 = 3.42, n_2 = 3.4187, $2a_o$ = 5μm, $2a_{fin}$ = 70μm, L = 500μm, λ_o = 1.55μm): solid line: output field calculated with the HGCM (M = 99, five longitudinal sections were used with w_o = 0.9 - 1.2 - 1.8 - 2.6 - 3.0μm, respectively); broken-dotted line: output field calculated with the DMM; dotted line: (local) fundamental mode at the broad end of the taper; broken line: free-space diffraction of the same initial field (as in Fig. 5.2.1) 120

5.2.11 Parabolic taper, the width of the taper varies along z as $a^2(z) = a^2_o + pz$, where p is the tapering parameter ($2a_o$ is the input width, $2a_{fin}$ is the output width, L is the length of the taper). 121

5.2.12 Field propagation in a) parabolic and b) linear tapered structures ($2a_o$ = 3μm, $2a_{fin}$ = 20μm, L = 500μm, n_1 = 3.33, n_2 = 3.32, λ = 0.86μm). solid line: SG (q = 16); dotted line: ST broken line: DMM. Parameters used for the HGCM

(M = 99, Δz = 5μm): parabolic taper: 5 longitudinal sections w_o = 0.3 - 0.5 - 0.7 - 0.9 - 1.1 μm, with $z_{section}$ = 40 - 80 - 120 - 200 - 500 μm; linear taper: 3 longitudinal sections w_o = 0.9 - 1.1 - 1.3 μm, with $z_{section}$ = 100 - 300 - 500 μm 123

5.2.13 Phase front at the output facet of a) parabolic, b) linear taper: solid line: SG (q = 16); dotted line: ST Linear taper: broken line: DMM. (Parameters in Fig. 5.2.12) 124

5.2.14 Far field from the output facet of a) parabolic, b) linear taper: solid line: SG (q = 16); dotted line: ST Linear taper: broken line: DMM. For the parameters refer to Fig. 5.2.12. 124

5.2.15 Convergence properties of the HGCM with respect to the parameter M, using a ST refractive index distribution: dotted line input field, broken-dotted line: M = 41, dotted line: M = 71, solid line: M = 99. Parameters for HGCM: M = 99, 3 longitudinal sections with $z_{section}$ = 100 - 300 - 500 μm, Δz = 5μm (for M = 41: w_o = 0.7 - 1.2 - 1.8μm; for M = 71: w_o = 0.7 - 1.2 - 1.5μm; for M = 99 : w_o = 0.4 - 0.9 - 1.2μm). Structure as in Fig. 5.2.12 . 125

5.2.16 Convergence properties of the HGCM with respect to the parameter M, using a SG refractive index distribution: dotted line input field, broken-dotted line M = 41, dotted line M = 71, solid line M = 99. (Parameters in Fig. 5.2.17.) 126

6.1.1 Tapered geometry device: a) top view: the shaded region defines the contact for current injection; b) front view: d s the thickness of the active layer 132

6.2.1 Iteration scheme for the solution of the nonlinear diffusion equation . 135

6.3.1 Linear diffusion equation in an unbounded region (contact width = 4μm, w_o = 1.05μm, D = 3.5 10^9 μm²/s, τ = 8ns, d = 0.1μm, J_o = 3.0 10^{-5} A/μm²): solid line: HGCM (M = 99), broken line: analytic solution. . . 137

6.3.2 Linear diffusion equation solved in a finite region (D = 3.5 10^9 μm²/s, τ = 8ns, d = 0.1μm, J_o = 3.0 10^{-5} A/μm², contact width = 10μm, x_o = −7μm, N_o = 3.0 10^6 μm^{-3} and x_3=7μm, N_1 = 1.0 10^6 μm^{-3}): solid line: HGCM (M = 99 , w_o = 1.05μm), broken line: analytic solution . . 138

6.3.3 Structure used for the solution of the linear diffusion equation with mixed boundary conditions (case 3). The current injection is limited to the region (x_1, x_2). The region $x < x_o$ is characterised by the parameters D_1, τ_1; the region $x_o < x < x_3$ by D_2, τ_2; and the region $x > x_3$ by D_3, τ_3 . . . 139

6.3.4 Linear diffusion equation solved in a finite region with mixed boundary conditions (D= 3.5 10^9 μm²/s, τ = 8ns, d = 0.1μm, J_o = 3.0 10^{-5} A/μm², contact = 19.5μm, v_o = 1.0 10^{10}μm/s, x_o = x_3 = −13.5μm, N_o = N_3 = 0μm^{-3}): solid line: HGCM (M = 99, w_o = 1.0μm), broken line: analytic solution 140

6.3.5 Solution of the nonlinear diffusion equation: solid line: HGCM (M = 99, w_o = 1.5μm at narrow end, w_o = 6.0μm at broad end), broken line: JTM. (Contact width at narrow end = 4μm, contact width at wide end = 100μm, D = 1.5 10^9 μm²/s, B_r = 10^2 $\mu m^3/s$, n_o = 10^4 μm^{-3}, J(x) = J_o inside the rib and J_o = 3.0 10^{-5} A / μm^2, thickness of active layer d = 0.17μm.) . 141

6.3.6 Carrier profile calculated including the effect of Auger recombination (γ = 1.0 10^{-5} μm^4, contact width = 4μm, D = 1.5 10^9 μm²/s, B_r = 10^2 $\mu m^3/s$, n_o = 10^4 μm^{-3}, J(x) = J_o inside the rib and J_o = 3.0 10^{-5} A/μm², and thickness of active layer d = 0.17 μm): solid line: HGCM (M = 99, w_o = 1.5μm); broken line: JTM 141

6.4.1 Effect of w_o parameter on the solution of the diffusion equation with the HGCM (M = 99): broken line: w_o = 0.5μm; broken-dotted line: w_o = 0.75μm; solid line: w_o = 2.0μm (contact width = 4μm, D = 1.5 10^9 $\mu m^2/s$, B_r = 10^2 $\mu m^3/s$, n_o = 10^4 μm^{-3}, J_o = 3.0 10^{-5} $A/\mu m^2$, d = 0.17 μm). . . 143

6.4.2 Peak value of the carrier distribution as a function of the width parameter used in the HGCM (M = 99). The variation is less than ±2%, (contact width = 4μm, D = 1.5 10^9 $\mu m^2/s$, B_r = 10^2 $\mu m^3/s$, n_o = 10^4 μm^{-3}, J_o = 3.0 10^{-5} $A/\mu m^2$, d = 0.17μm) 143

6.4.3 Carrier profile calculated with the HGCM with varying M: broken line: M = 31; broken-dotted line: M = 61; solid line: M = 111 (contact width = 4μm, D = 1.5 10^9 $\mu m^2/s$, B_r = 10^2 $\mu m^3/s$, n_o = 10^4 μm^{-3}, $J_o = 3.0\ 10^{-5}\ A/\mu m^2$, d = 0.17 μm). 144

6.4.4 Dependence of the peak value of the carrier distribution profile with varying M. Note that the variation is less than ±2% (contact width = 4μm, $D = 1.5\ 10^9 \mu m^2/s$, B_r = 10^2 $\mu m^3/s$, n_o = 10^4 μm^{-3}, J_o = 3.0 10^{-5} $A/\mu m^2$, d = 0.17μm) 145

7.1.1 Towards tapered geometry devices: a) ordinary stripe laser, b) multi-stripe laser array, c) broad area laser, d) tapered geometry laser and bow-tie laser 156

7.2.1 a) Tapered device; c) front view: d is the thickness of the active layer; b) top-plane view: $2a_o$ and $2a_{fin}$ are the widths of narrower and wider facets, respectively, L is the length of the device. The tapered shaded region also corresponds to the area of the metal contact. . . . 157

7.2.2 Schematic for the iteration procedure used with the self-consistent HGCM model for a) travelling-wave amplifiers, b) laser devices; (gp indicates the stimulated emission term in the carrier diffusion equation, equation (7.2.14); $\varepsilon_{(j)}(x, z)$ the complex dielectric distribution calculated in the j-th iteration; Δz the longitudinal step) 164

7.3.1 Linearly tapered travelling-wave amplifier, near field (z = L): solid line: HGCM; broken line: DMM; dotted-broken line: input field; dotted line: fundamental mode at z = L ($J_o = 1.5\ 10^{-5}$ A/μm², P_i = 5mW, P_{out} = 32mW) 166

7.3.2 Tapered amplifier: single pass gain for $J_o = 1.5\ 10^{-5}$ A/μm², and input power P_i = 5mW: solid line: HGCM (P_{out} = 32mW); broken line: DMM (P_{out} = 28mW) 167

7.3.3 Light-current characteristics for the tapered geometry travelling wave-amplifier: broken line: P_{in} = 1mW; solid line: P_{in} = 5mW; dotted line: DMM with P_{in} = 5mW 168

7.3.4 Field propagation in the tapered travelling-wave amplifier: a) forward, b) reverse travelling field ($J_o = 1.0\ 10^{-5}$ A/μm², P_{in} = 0.1mW) . . 170

7.3.5 Field propagation in the tapered travelling-wave amplifier: comparison of the fields obtained in Fig. 7.3.4 a) and b) at z = 0. . . . 171

7.3.6 Near field for tapered laser at z = 0 (broken line), and at z = L (continuous line). 173

7.3.7 Output power after each iteration 173

7.3.8 Carrier density distribution in the tapered laser: broken line: in the absence of the optical field in the cavity; continuous line: self-consistent solution 174

7.3.9 Near field for the tapered laser at z = 0 (broken line) and at z = L (continuous line). 175

7.3.10 Output power obtained after each round-trip iteration 175

7.3.11 Tapered laser: a) near field at z = 0 (broken line) and at z = L (continuous line); b) phase front at z = L; c) far field intensity profile. . . 177

7.3.12 Tapered laser: broken line: without the optical field in the cavity; continuous line: self-consistent solution. 178

7.3.13 Tapered laser: output power obtained after each round-trip iteration: (+) J_o = 1.0 10^{-5} A/μm², (◦) J_o = 1.2 10^{-5} A/μm², (•) J_o = 1.5 10^{-5} A/μm². . 179

7.3.14 Near field for the tapered laser: broken line: near field at z = 0; continuous line: near field at z = L. 180

7.3.15 Model for the inclusion of deflectors (darker shaded region at z = 0) in the tapered laser. 181

7.3.16 Power as a function of the iteration number for the tapered laser with deflectors. 182

7.3.17 Near field distribution: broken line: at z = 0; continuous line: at z = L 183

7.3.18 Carrier distribution at the narrow and wide ends of the tapered laser: broken line: without the optical filed in the cavity; continuous line: self-consistent solution. 183

7.4.1 Parabolically tapered laser: a) near field: dotted line: at z = 0; solid line: at z = L; b) phase front at z = L; c) far field intensity profile. In all three graphs: broken line: DMM. 186-187

7.4.2 Parabolically tapered laser: output power obtained after each round trip iteration: (+) $J_o = 1.0\ 10^{-5}$ A/μm², (◦) $J_o = 1.2\ 10^{-5}$ A/μm², (•) $J_o = 1.5\ 10^{-5}$ A/μm².. 187

7.4.3 Parabolically tapered laser: broken line: without the optical field in the cavity; continuous line: self-consistent solution. 188

7.5.1 Stripe laser geometry. 189

7.5.2 Near field: comparison between the gain-guided (broken line) and the index-guided stripe laser (continuous line). 191

7.5.3 Power calculated at each iteration for different injection current densities: a) current densities used for the gain-guided device: (o) $J_o = 8.0\ 10^{-5}$ A/μm², (•) $J_o = 9.0\ 10^{-5}$ A/μm²; current densities used for the index-guided device b): (o) $J_o = 5.0\ 10^{-5}$ A/μm², (•) $J_o = 8.0\ 10^{-5}$ A/μm² 191

7.5.4 Refrective index distribution and carrier density for gain-guided and index-guided laser: distribution before (broken line) and after (continuous line) self-consistency. 192

7.5.5 Evolution of the near field in the iterative process: broken line at z = 0; continuous line at z = L. Note that in a) and b) the scale is orders of magnitude smaller than that used in c) and d). . . . 194

A7.1-1 Variation of the (calculated) effective refractive index and of the confinement factor along the vertical (y) direction, with respect to the etching depth; a) Quantum Well material (QT829A), b) bulk material (CB539) . 200

A7.1-1 Schematic of a typical ridge waveguide semiconductor device; $2a_o$ is the ribwidth 201